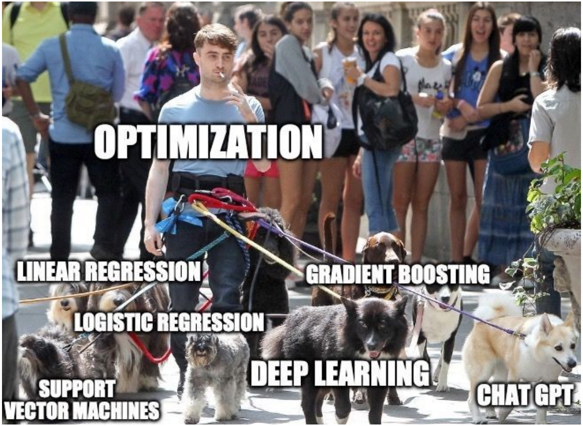

Introduction to Optimization

K. R. Sahasranand

Data Science

sahasranand@iitpkd.ac.in

Optimization thus far

minimize (or maximize) –

- Specific functions **without** constraints   Least squares

Optimization thus far

minimize (or maximize) –

- Specific functions **without** constraints — Least squares
- Specific functions subject to **linear** constraints — Minimum ℓ_2 norm

Optimization thus far

minimize (or maximize) –

- ∗ Specific functions without constraints — Least squares
- ∗ Specific functions subject to linear constraints — Minimum ℓ_2 norm
- ∗ Linear functions subject to linear constraints — LP

Optimization thus far

minimize (or maximize) –

- ∗ Specific functions without constraints — Least squares
- ∗ Specific functions subject to linear constraints — Minimum ℓ_2 norm
- ∗ Linear functions subject to linear constraints — LP
- ∗ Nonlinear $\mathbb{R} \to \mathbb{R}$ functions without constraints — Line search

Optimization thus far

minimize (or maximize) –

- $*$ Specific functions without constraints — Least squares
- $*$ Specific functions subject to linear constraints — Minimum ℓ_2 norm
- $*$ Linear functions subject to linear constraints — LP
- $*$ Nonlinear $\mathbb{R} \to \mathbb{R}$ functions without constraints — Line search
- $*$ Nonlinear $\mathbb{R}^n \to \mathbb{R}$ functions without constraints — Gradient methods

Optimization thus far

minimize (or maximize) –

- $*$ Specific functions without constraints — Least squares
- $*$ Specific functions subject to linear constraints — Minimum ℓ_2 norm
- $*$ Linear functions subject to linear constraints — LP
- $*$ Nonlinear $\mathbb{R} \to \mathbb{R}$ functions without constraints — Line search
- $*$ Nonlinear $\mathbb{R}^n \to \mathbb{R}$ functions without constraints — Gradient methods
- $*$ Nonlinear $\mathbb{R}^n \to \mathbb{R}$ functions subject to general constraints
  - *Equality constraints* — Lagrange multipliers
  - *Inequality constraints* — Karush-Kuhn-Tucker conditions

Problems with equality constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \qquad i = 1, \dots, m \\ \text{where} \quad & x \in \mathbb{R}^n, \\ & f : \mathbb{R}^n \to \mathbb{R}, \\ & h_i : \mathbb{R}^n \to \mathbb{R}, \quad m \leq n. \end{aligned}$$

First-order necessary condition

Lagrange's theorem for $n = 2, m = 1$

Theorem – Let x^* be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint

$$h(x) = 0, h : \mathbb{R}^2 \to \mathbb{R}.$$

Then, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel. That is, if

$$\nabla h(x^*) \neq 0,$$

then there exists a scalar λ such that

$$\nabla f(x^*) + \lambda \nabla h(x^*) = 0.$$

First-order necessary condition

Lagrange's theorem for $n = 2, m = 1$

Theorem – Let x^* be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint

$$h(x) = 0, h : \mathbb{R}^2 \to \mathbb{R}.$$

Then, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel. That is, if

$$\nabla h(x^*) \neq 0,$$

then there exists a scalar λ such that

$$\nabla f(x^*) + \lambda \nabla h(x^*) = 0.$$

$\boxed{\lambda \sim \text{“Lagrange multiplier”}}$

First-order necessary condition

Lagrange's theorem for $n = 2, m = 1$

Theorem – Let x^* be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint

$$h(x) = 0, h : \mathbb{R}^2 \to \mathbb{R}.$$

Then, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel. That is, if

$$\nabla h(x^*) \neq 0,$$

then there exists a scalar λ such that

$$\nabla f(x^*) + \lambda \nabla h(x^*) = 0.$$

$\boxed{\lambda \sim \text{“Lagrange multiplier”}}$

Joseph-Louis Lagrange

First-order necessary condition

Lagrange's theorem for $n = 2, m = 1$

Theorem – Let x^* be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ subject to the constraint

$$h(x) = 0, h : \mathbb{R}^2 \to \mathbb{R}.$$

Then, $\nabla f(x^*)$ and $\nabla h(x^*)$ are parallel. That is, if

$$\nabla h(x^*) \neq 0,$$

then there exists a scalar λ such that

$$\nabla f(x^*) + \lambda \nabla h(x^*) = 0.$$

$\boxed{\lambda \sim \text{“Lagrange multiplier”}}$

Joseph-Louis Lagrange

Why? Illustrate on board.

An example

$$\begin{aligned} \text{minimize } & x_1 x_2 + 1 \\ \text{subject to } & x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

An example

$$\begin{aligned} \text{minimize } & x_1 x_2 + 1 \\ \text{subject to } & x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

We have

$$\nabla f(x) = \begin{bmatrix} x_2 & x_1 \end{bmatrix}^\top \qquad ; \qquad \nabla h(x) = \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}^\top$$

An example

$$\begin{aligned} \text{minimize } & x_1 x_2 + 1 \\ \text{subject to } & x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

We have

$$\nabla f(x) = \begin{bmatrix} x_2 & x_1 \end{bmatrix}^\top \qquad ; \qquad \nabla h(x) = \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}^\top$$

Therefore,

$$\begin{aligned} x_2 + \lambda \cdot 2x_1 &= 0 \\ x_1 + \lambda \cdot 2x_2 &= 0 \\ x_1^2 + x_2^2 &= 1 \end{aligned}$$

An example

$$\begin{aligned} \text{minimize } & x_1 x_2 + 1 \\ \text{subject to } & x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

We have

$$\nabla f(x) = \begin{bmatrix} x_2 & x_1 \end{bmatrix}^\top \qquad ; \qquad \nabla h(x) = \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}^\top$$

Therefore,

$$\begin{aligned} x_2 + \lambda \cdot 2x_1 &= 0 \\ x_1 + \lambda \cdot 2x_2 &= 0 \\ x_1^2 + x_2^2 &= 1 \end{aligned}$$

(complete on board)

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

$$\boxed{\text{maximize } x_1x_2x_3 \text{ subject to } x_1x_2 + x_2x_3 + x_3x_1 = A/2}$$

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

$$\boxed{\text{maximize } x_1x_2x_3 \text{ subject to } x_1x_2 + x_2x_3 + x_3x_1 = A/2}$$

(Exer: complete the calculation)

Ans: $x_1^* = x_2^* = x_3^* = \sqrt{A/6}$.

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

$$\boxed{\text{maximize } x_1x_2x_3 \text{ subject to } x_1x_2 + x_2x_3 + x_3x_1 = A/2}$$

(Exer: complete the calculation)

Ans: $x_1^* = x_2^* = x_3^* = \sqrt{A/6}$.

· Is x^* the maximum or the minimum?

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

$$\boxed{\text{maximize } x_1x_2x_3 \text{ subject to } x_1x_2 + x_2x_3 + x_3x_1 = A/2}$$

(Exer: complete the calculation)

Ans: $x_1^* = x_2^* = x_3^* = \sqrt{A/6}$.

- Is x^* the maximum or the minimum?
- Consider $\widetilde{x} = \begin{bmatrix} 1 & 1 & \frac{A-2}{4} \end{bmatrix}^\top$ Is $\widetilde{x}$ feasible?

Another example

Given a fixed area of cardboard, say $A\ m^2$, we wish to construct a closed (cuboidal) box with the maximum volume.

$$\boxed{\text{maximize } x_1x_2x_3 \text{ subject to } x_1x_2 + x_2x_3 + x_3x_1 = A/2}$$

(Exer: complete the calculation)

Ans: $x_1^* = x_2^* = x_3^* = \sqrt{A/6}$.

- Is x^* the maximum or the minimum?
- Consider $\widetilde{x} = \begin{bmatrix} 1 & 1 & \frac{A-2}{4} \end{bmatrix}^\top$ Is $\widetilde{x}$ feasible?
- Is $\widetilde{x}_1\widetilde{x}_2\widetilde{x}_3 < x_1^*x_2^*x_3^*$?

Lagrange's theorem

General case

Definition – A point x^* satisfying the constraints

$$h_1(x^*) = 0, \cdots, h_m(x^*) = 0$$

is said to be a regular point of the constraints if the gradient vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent.

Lagrange's theorem

General case

Definition – A point x^* satisfying the constraints

$$h_1(x^*) = 0, \cdots, h_m(x^*) = 0$$

is said to be a regular point of the constraints if the gradient vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent.

Example: $n = 3, m = 2$ and $h_1(x) = x_1 \; ; \; h_2(x) = x_2 - x_3^2$.

Lagrange's theorem

General case

Lagrange's theorem – Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, there exists $\lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Lagrange's theorem
General case

Lagrange's theorem– Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \;; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, $\exists \lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Introduce a Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^\top h(x).$$

Lagrange's theorem

General case

Lagrange's theorem– Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, $\exists \lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Introduce a Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^\top h(x).$$

Lagrange condition $\equiv$ FONC for unconstrained optimization for $\mathcal{L}$

Lagrange's theorem

General case

Lagrange's theorem– Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \ ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, $\exists \lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Introduce a Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^\top h(x).$$

Lagrange condition $\equiv$ FONC for unconstrained optimization for $\mathcal{L}$

$$D\mathcal{L}(x, \lambda) = \begin{bmatrix} D_x\mathcal{L}(x, \lambda) & D_\lambda\mathcal{L}(x, \lambda) \end{bmatrix}^\top = 0^\top.$$

Lagrange's theorem

General case

Lagrange's theorem– Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \ ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, $\exists \lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Introduce a Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^\top h(x).$$

Lagrange condition $\equiv$ FONC for unconstrained optimization for $\mathcal{L}$

$$D\mathcal{L}(x, \lambda) = \begin{bmatrix} D_x\mathcal{L}(x, \lambda) & D_\lambda\mathcal{L}(x, \lambda) \end{bmatrix}^\top = 0^\top.$$

That is,

$$\begin{aligned} D_x\mathcal{L}(x, \lambda) &= Df(x) + \lambda^\top Dh(x) &= 0^\top \qquad &\text{Lagrange's theorem} \\ D_\lambda\mathcal{L}(x, \lambda) &= h(x) &= 0^\top \qquad &\text{Constraints} \end{aligned}$$

Lagrange's theorem

General case

Lagrange's theorem– Let x^* be a local minimizer (or maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \le n.$$

Assuming that x^* is a regular point, $\exists \lambda \in \mathbb{R}^m$ such that

$$Df(x^*) + \lambda^\top Dh(x^*) = 0^\top.$$

Introduce a Lagrangian function

$$\mathcal{L}(x, \lambda) := f(x) + \lambda^\top h(x).$$

FONC on (unconstrained) $\mathcal{L}$ yields

$$D_x\mathcal{L}(x,\lambda) = Df(x) + \lambda^\top Dh(x) \qquad = 0^\top \qquad \text{Lagrange's theorem}$$
$$D_\lambda\mathcal{L}(x,\lambda) = h(x) \qquad = 0^\top \qquad \text{Constraints}$$

$n + m$ equations in $n + m$ unknowns

An example

$n = 3$ variables, $m = 2$ constraints

$$\begin{aligned} \text{maximize } & 4x_1 + 6x_2 - 2x_3 \\ \text{subject to } & 2x_1 + x_2 + 5x_3 = 1 \\ \text{and } & x_2^2 + 2x_3^3 = 22. \end{aligned}$$

An example

$n = 3$ variables, $m = 2$ constraints

$$\text{maximize } 4x_1 + 6x_2 - 2x_3$$
$$\text{subject to } 2x_1 + x_2 + 5x_3 = 1$$
$$\text{and } x_2^2 + 2x_3^3 = 22.$$

By Lagrange's theorem,

$$\nabla f(x) = \lambda_1 \nabla h_1(x) + \lambda_2 \nabla h_2(x)$$

An example

$n = 3$ variables, $m = 2$ constraints

$$\text{maximize } 4x_1 + 6x_2 - 2x_3$$
$$\text{subject to } 2x_1 + x_2 + 5x_3 = 1$$
$$\text{and } x_2^2 + 2x_3^3 = 22.$$

By Lagrange's theorem,

$$\nabla f(x) = \lambda_1 \nabla h_1(x) + \lambda_2 \nabla h_2(x)$$

where

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} 4 & 6 & -2 \end{bmatrix}^\top \\ \nabla h_1(x) &= \begin{bmatrix} 2 & 1 & 5 \end{bmatrix}^\top \\ \nabla h_2(x) &= \begin{bmatrix} 0 & 2x_2 & 4x_3 \end{bmatrix}^\top \end{aligned}$$

An example

$n = 3$ variables, $m = 2$ constraints

$$\begin{aligned} \text{maximize } & 4x_1 + 6x_2 - 2x_3 \\ \text{subject to } & 2x_1 + x_2 + 5x_3 = 1 \\ \text{and } & x_2^2 + 2x_3^3 = 22. \end{aligned}$$

By Lagrange's theorem,

$$\nabla f(x) = \lambda_1 \nabla h_1(x) + \lambda_2 \nabla h_2(x)$$

where

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} 4 & 6 & -2 \end{bmatrix}^\top \\ \nabla h_1(x) &= \begin{bmatrix} 2 & 1 & 5 \end{bmatrix}^\top \\ \nabla h_2(x) &= \begin{bmatrix} 0 & 2x_2 & 4x_3 \end{bmatrix}^\top \end{aligned}$$

(complete on board)

Second-order necessary conditions

Let x^* be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, there exists $\lambda \in \mathbb{R}^m$ such that

i. $Df(x^*) + \lambda^\top Dh(x^*) = 0^\top$

where

$T(x^*) = \{y : Dh(x^*)y = 0\}$ and
$L(x^*, \lambda)$ is the Hessian of $\mathcal{L}(x^*, \lambda)$.

Second-order necessary conditions

Let x^* be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, there exists $\lambda \in \mathbb{R}^m$ such that

i. $Df(x^*) + \lambda^\top Dh(x^*) = 0^\top$

ii. For all $y \in T(x^*)$, $y^\top L(x^*, \lambda) y \geq 0$.

where

$T(x^*) = \{y : Dh(x^*)y = 0\}$ and

$L(x^*, \lambda)$ is the Hessian of $\mathcal{L}(x^*, \lambda)$.

Second-order necessary conditions

Let x^* be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, there exists $\lambda \in \mathbb{R}^m$ such that

i. $Df(x^*) + \lambda^\top Dh(x^*) = 0^\top$

ii. For all $y \in T(x^*)$, $y^\top L(x^*, \lambda) y \geq 0$.

where

$T(x^*) = \{y : Dh(x^*)y = 0\}$ and
$L(x^*, \lambda)$ is the Hessian of $\mathcal{L}(x^*, \lambda)$.

That is,

$$L(x, \lambda) = F(x) + \lambda_1 H_1(x) + \cdots + \lambda_m H_m(x).$$

Second-order necessary conditions

Let x^* be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to

$$h(x) = 0 \; ; \qquad h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n.$$

Assuming that x^* is a regular point, there exists $\lambda \in \mathbb{R}^m$ such that

i. $Df(x^*) + \lambda^\top Dh(x^*) = 0^\top$

ii. For all $y \in T(x^*)$, $y^\top L(x^*, \lambda) y \geq 0$.

where

$T(x^*) = \{y : Dh(x^*)y = 0\}$ and

$L(x^*, \lambda)$ is the Hessian of $\mathcal{L}(x^*, \lambda)$.

Converse: **Second-order sufficient conditions**

Second-order sufficient conditions

Theorem – Suppose that $f, h \in \mathcal{C}^2$ and there exists a point $x^* \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^m$ such that:

i. $Df(x^*) + \lambda^\top Dh(x^*) = 0^\top$

ii. For all $y \in T(x^*), y^\top L(x^*, \lambda) y > 0$.

Then, x^* is a strict local minimizer of f subject to $h(x) = 0$.

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

where $Q > 0, A \in \mathbb{R}^{m\times n}, m < n,$ rank $A = m.$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

$$\text{where } Q > 0, A \in \mathbb{R}^{m\times n}, m < n, \text{rank } A = m.$$

Lagrangian function: $\mathcal{L}(x,\lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

$$\text{where } Q > 0, A \in \mathbb{R}^{m\times n}, m < n, \text{rank } A = m.$$

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

Lagrange condition: $D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Q x \text{ subject to } Ax = b$$

$$\text{where } Q > 0, A \in \mathbb{R}^{m \times n}, m < n, \text{rank } A = m.$$

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Q x + \lambda^\top (b - Ax)$.

Lagrange condition: $D_x \mathcal{L}(x^*, \lambda) = x^{*\top} Q - \lambda^{*\top} A = 0^\top$

rewriting which yields

$$x^* = Q^{-1} A^\top \lambda$$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

where $Q > 0, A \in \mathbb{R}^{m\times n}, m < n,$ rank $A = m.$

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax).$

Lagrange condition: $D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

where $Q > 0, A \in \mathbb{R}^{m\times n}, m < n,$ rank $A = m$.

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

Lagrange condition: $D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

Since $Ax^* = b$ and since $AQ^{-1}A^\top$ is invertible (why?),

$$\lambda = (AQ^{-1}A^\top)^{-1}b$$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

where $Q > 0, A \in \mathbb{R}^{m\times n}, m < n,$ rank $A = m$.

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

Lagrange condition: $D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

Since $Ax^* = b$ and since $AQ^{-1}A^\top$ is invertible (why?),

$$\lambda = (AQ^{-1}A^\top)^{-1}b$$

whereby

$$\boxed{x^* = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b}$$

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

$$\text{where } Q > 0, A \in \mathbb{R}^{m\times n}, m < n, \text{rank } A = m.$$

Lagrangian function: $\mathcal{L}(x, \lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

Lagrange condition: $D_x\mathcal{L}(x^*, \lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

Since $Ax^* = b$ and since $AQ^{-1}A^\top$ is invertible (why?),

$$\lambda = (AQ^{-1}A^\top)^{-1}b$$

whereby

$$\boxed{x^* = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b}$$

Is x^* the minimum or the maximum?

An example

Quadratic programming

$$\text{minimize } \frac{1}{2}x^\top Qx \text{ subject to } Ax = b$$

$$\text{where } Q > 0, A \in \mathbb{R}^{m\times n}, m < n, \text{rank } A = m.$$

Lagrangian function: $\mathcal{L}(x,\lambda) = \frac{1}{2}x^\top Qx + \lambda^\top(b - Ax)$.

Lagrange condition: $D_x\mathcal{L}(x^*,\lambda) = x^{*\top}Q - \lambda^{*\top}A = 0^\top$

rewriting which yields

$$x^* = Q^{-1}A^\top\lambda \qquad \Longrightarrow \qquad Ax^* = AQ^{-1}A^\top\lambda.$$

Since $Ax^* = b$ and since $AQ^{-1}A^\top$ is invertible (why?),

$$\lambda = (AQ^{-1}A^\top)^{-1}b$$

whereby

$$\boxed{x^* = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b}$$

Is x^* the minimum or the maximum? $L(x^*,\lambda) = Q > 0.$

Problems with inequality constraints

$$\text{minimize} \quad f(x) \qquad \text{subject to}$$
$$h(x) = 0,$$
$$g(x) \leq 0,$$

where

$$f : \mathbb{R}^n \to \mathbb{R}, \quad h : \mathbb{R}^n \to \mathbb{R}^m; \; m \leq n, \quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Problems with inequality constraints

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R}, \quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n, \quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

The inequality constraint:

$g_j(x) \leq 0$	active at x^*	if $g_j(x^*) = 0$.
$g_j(x) \leq 0$	inactive at x^*	if $g_j(x^*) < 0$.

Problems with inequality constraints

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

The inequality constraint:

$g_j(x) \leq 0$	active at x^*	if $g_j(x^*) = 0$.
$g_j(x) \leq 0$	inactive at x^*	if $g_j(x^*) < 0$.

Let x^* satisfy $h(x^*) = 0$ and $g(x^*) \leq 0$, and

$$J(x^*) = \{j : g_j(x^*) = 0\}$$ "active constraints"

Then, we say that x^* is a **regular point** if the vectors

$$\nabla h_i(x^*), \nabla g_j(x^*), \quad 1 \leq i \leq m, j \in J(x^*)$$

are linearly independent.

Karush-Kuhn-Tucker conditions

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \leq 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \geq 0$

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$

iii. $\mu^\top g(x^*) = 0.$

Karush-Kuhn-Tucker conditions

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \leq 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \geq 0$
ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$
iii. $\mu^\top g(x^*) = 0$.

The constraints in the problem are

iv. $h(x^*) = 0$
v. $g(x^*) \leq 0$.

Karush-Kuhn-Tucker conditions

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \le 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R}, \quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \le n, \quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \le 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \ge 0$

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$

iii. $\mu^\top g(x^*) = 0.$

The constraints in the problem are

iv. $h(x^*) = 0$

v. $g(x^*) \le 0.$

Solve for x^*, λ, μ using the above.

Karush-Kuhn-Tucker conditions

minimize $f(x)$ subject to $h(x) = 0,\ g(x) \leq 0,$

where $f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \leq 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \geq 0$

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$ Stationarity

iii. $\mu^\top g(x^*) = 0$.

The constraints in the problem are

iv. $h(x^*) = 0$

v. $g(x^*) \leq 0$.

Solve for x^*, λ, μ using the above.

Karush-Kuhn-Tucker conditions

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \le 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \le n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \le 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \ge 0$

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$ — Stationarity

iii. $\mu^\top g(x^*) = 0.$

The constraints in the problem are

iv. $h(x^*) = 0$ — Primal feasibility

v. $g(x^*) \le 0.$ — Primal feasibility

Solve for x^*, λ, μ using the above.

Karush-Kuhn-Tucker conditions

minimize $f(x)$ subject to $h(x) = 0,\ g(x) \leq 0,$

where $f : \mathbb{R}^n \to \mathbb{R}, \quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n, \quad g : \mathbb{R}^n \to \mathbb{R}^p.$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \leq 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \geq 0$ — Dual feasibility

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$ — Stationarity

iii. $\mu^\top g(x^*) = 0.$

The constraints in the problem are

iv. $h(x^*) = 0$ — Primal feasibility

v. $g(x^*) \leq 0.$ — Primal feasibility

Solve for x^*, λ, μ using the above.

Karush-Kuhn-Tucker conditions

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ g(x) \leq 0,$$
$$\text{where } f : \mathbb{R}^n \to \mathbb{R},\quad h : \mathbb{R}^n \to \mathbb{R}^m;\ m \leq n,\quad g : \mathbb{R}^n \to \mathbb{R}^p.$$

Theorem – Let $f, g, h \in \mathcal{C}$. Let x^* be a regular point and a local minimizer for the problem

$$\text{minimize } f(x) \ \text{ subject to } h(x) = 0,\ g(x) \leq 0.$$

Then, there exists $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ such that

i. $\mu \geq 0$ — Dual feasibility

ii. $Df(x^*) + \lambda^\top Dh(x^*) + \mu^\top Dg(x^*) = 0^\top$ — Stationarity

iii. $\mu^\top g(x^*) = 0.$ — Complementary slackness

The constraints in the problem are

iv. $h(x^*) = 0$ — Primal feasibility

v. $g(x^*) \leq 0.$ — Primal feasibility

Solve for x^*, λ, μ using the above.

An example

Communication over parallel channels $i = 1, \ldots, k$

- The communication rate over channel i is given by

$$\frac{1}{2} \log \left(1 + \frac{x_i}{\sigma_i^2} \right)$$

where σ_i^2 is the noise power and x_i denotes the signal power.

An example

Communication over parallel channels $i = 1, \ldots, k$

- The communication rate over channel i is given by

$$\frac{1}{2} \log \left(1 + \frac{x_i}{\sigma_i^2} \right)$$

  where σ_i^2 is the noise power and x_i denotes the signal power.

- The goal is to maximize the sum of rates over all channels.

An example

Communication over parallel channels $i = 1, \ldots, k$

- The communication rate over channel i is given by

$$\frac{1}{2} \log \left(1 + \frac{x_i}{\sigma_i^2} \right)$$

  where σ_i^2 is the noise power and x_i denotes the signal power.

- The goal is to maximize the sum of rates over all channels.

- The noise power values are given and we need to choose the signal power values subject to the constraint that the total power expended does not exceed a given value $P > 0$.

An example

Communication over parallel channels $i = 1, \ldots, k$

- The communication rate over channel i is given by

$$\frac{1}{2} \log\left(1 + \frac{x_i}{\sigma_i^2}\right)$$

  where σ_i^2 is the noise power and x_i denotes the signal power.

- The goal is to maximize the sum of rates over all channels.

- The noise power values are given and we need to choose the signal power values subject to the constraint that the total power expended does not exceed a given value $P > 0$.

- Indeed, $x_i \geq 0$ for all $i = 1, \ldots, n$.

An example

Communication over parallel channels $i = 1, \dots, k$

- The communication rate over channel i is given by

$$\frac{1}{2}\log\left(1 + \frac{x_i}{\sigma_i^2}\right)$$

where σ_i^2 is the noise power and x_i denotes the signal power.

- The goal is to maximize the sum of rates over all channels.

- The noise power values are given and we need to choose the signal power values subject to the constraint that the total power expended does not exceed a given value $P > 0$.

- Indeed, $x_i \geq 0$ for all $i = 1, \dots, n$.

$$\max_{x_1,\dots,x_k} \frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) \text{ subject to } \begin{cases} \sum_{i=1}^{k} x_i \leq P, \\ x_i \geq 0. \end{cases}$$

An example

Communication over parallel channels $i = 1, \ldots, k$

$$\max_{x_1,\ldots,x_k} \frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) \text{ subject to } \begin{cases} \sum_{i=1}^{k} x_i \leq P, \\ x_i \geq 0. \end{cases}$$

An example

Communication over parallel channels $i = 1, \ldots, k$

$$\max_{x_1,\ldots,x_k} \frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) \text{ subject to } \begin{cases} \sum_{i=1}^{k} x_i \leq P, \\ x_i \geq 0. \end{cases}$$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu \left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

An example

Communication over parallel channels $i = 1, \dots, k$

$$\max_{x_1,\dots,x_k} \frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) \text{ subject to } \begin{cases} \sum_{i=1}^{k} x_i \le P, \\ x_i \ge 0. \end{cases}$$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2} \sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu \left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \le P, \\ x_i^* \ge 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu \left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \ge 0, \\ \mu \ge 0 \end{cases}$

An example

Communication over parallel channels $i = 1, \ldots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i\left(-x_i - 0\right).$$

Stationarity: $\frac{\partial J}{\partial x_i}\Big|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

An example

Communication over parallel channels $i = 1, \dots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i\left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

$$\frac{\partial J}{\partial x_i} = -\frac{1}{2} \cdot \frac{1}{x_i + \sigma_i^2} + \mu - \mu_i = 0$$

An example

Communication over parallel channels $i = 1, \ldots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

$$\frac{\partial J}{\partial x_i} = -\frac{1}{2} \cdot \frac{1}{x_i + \sigma_i^2} + \mu - \mu_i = 0$$

which yields

$$2\mu = 2\mu_i + \frac{1}{x_i + \sigma_i^2}.$$

An example

Communication over parallel channels $i = 1, \dots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

$$\frac{\partial J}{\partial x_i} = -\frac{1}{2} \cdot \frac{1}{x_i + \sigma_i^2} + \mu - \mu_i = 0$$

which yields

$$2\mu = 2\mu_i + \frac{1}{x_i + \sigma_i^2}.$$

Case 1: $\boxed{\frac{1}{2\mu} \leq \sigma_i^2}$

Case 2: $\boxed{\frac{1}{2\mu} > \sigma_i^2}$

(continue on board)

An example

Communication over parallel channels $i = 1, \ldots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

The solution is given by

$$x_i^* = \begin{cases} 0 & \text{if } \frac{1}{2\mu} \leq \sigma_i^2 \\ \frac{1}{2\mu} - \sigma_i^2 & \text{otherwise.} \end{cases}$$

An example

Communication over parallel channels $i = 1, \ldots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i\left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

The solution is given by

$$x_i^* = \left(\frac{1}{2\mu} - \sigma_i^2\right)_+ \quad ; \qquad (a)_+ = \max\{a, 0\}.$$

An example

Communication over parallel channels $i = 1, \dots, k$

Lagrangian function:

$$J(x^n, \mu, \mu^n) = -\frac{1}{2}\sum_{i=1}^{k} \log\left(1 + \frac{x_i}{\sigma_i^2}\right) + \mu\left(\sum_{i=1}^{n} x_i - P\right) + \sum_{i=1}^{n} \mu_i \left(-x_i - 0\right).$$

Stationarity: $\left.\frac{\partial J}{\partial x_i}\right|_{x_i^*} = 0$

Primal feasibility: $\begin{cases} \sum_i x_i^* \leq P, \\ x_i^* \geq 0 \end{cases}$

Complementary slackness: $\begin{cases} \mu_i x_i^* = 0, \\ \mu\left(\sum_i x_i^* - P\right) = 0 \end{cases}$

Dual feasibility: $\begin{cases} \mu_i \geq 0, \\ \mu \geq 0 \end{cases}$

The solution is given by

$$x_i^* = \left(\frac{1}{2\mu} - \sigma_i^2\right)_+ ; \qquad (a)_+ = \max\{a, 0\}.$$

By complementary slackness, we choose μ such that

$$\sum_{i=1}^{n} x_i^* = \sum_{i=1}^{n} \left(\frac{1}{2\mu} - \sigma_i^2\right)_+ = P.$$

Thank You!