8. Constrained optimization

- equality constrained problems
- inequality constrained problems
- quadratic problems with linear constraints
- projected gradient descent
- penalty method

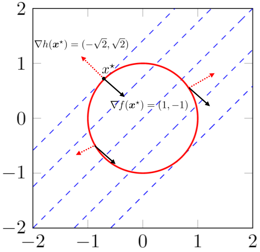
Equality constrained problems

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & h_i(\boldsymbol{x}) = \mathbf{0}, \quad i = 1, \ldots, p \end{array} \tag{8.1}$$

- $f : \mathbb{R}^n \to \mathbb{R}$
- $h_i : \mathbb{R}^n \to \mathbb{R}$
- we let $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_p(\boldsymbol{x}))$
- a point $\boldsymbol{x}$ satisfying $h(\boldsymbol{x}) = \mathbf{0}$ is called a *feasible point*

Example 8.1

minimize	$x_1 - x_2$
subject to	$x_1^2 + x_2^2 = 1$



- circle represent the constraint
- dotted lines are the level sets ($f(\boldsymbol{x}) = x_1 - x_2 = \gamma$) at different values of γ
- black arrows shows the direction of the gradient $\nabla f(\boldsymbol{x}) = (1, -1)$
- the global minimizer is $\boldsymbol{x}^\star = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$
- the gradients $\nabla f(\boldsymbol{x}^\star)$ and $\nabla h(\boldsymbol{x}^\star)$ are parallel (linearly dependent):

$$\nabla f(\boldsymbol{x}^\star) = -\lambda \nabla h(\boldsymbol{x}^\star)$$

where $\lambda = 1/\sqrt{2}$

Motivation of optimality conditions

suppose that we only have one constraint ($p = 1$) and consider the problem

$$\text{minimize } f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

where $\lambda \in \mathbb{R}$ is an adjustable parameter

- if there exists some $\lambda^\star$ such that the solution of the above problem, $\boldsymbol{x}^\star$, satisfies $h(\boldsymbol{x}^\star) = 0$, *i.e.*, there exists some $\lambda^\star$ such that:

  $$\nabla f(\boldsymbol{x}^\star) + \lambda^\star \nabla h(\boldsymbol{x}^\star) = \boldsymbol{0} \quad \text{and} \quad h(\boldsymbol{x}^\star) = 0$$

  then, we have

  $$f(\boldsymbol{x}^\star) = f(\boldsymbol{x}^\star) + \lambda^\star h(\boldsymbol{x}^\star) \leq f(\boldsymbol{x}) + \lambda^\star h(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x}$$

  hence, $f(\boldsymbol{x}^\star) \leq f(\boldsymbol{x})$ for all feasible $\boldsymbol{x}$ ($\boldsymbol{x}^\star$ is a solution to the original problem (8.1))

- we can transform the constrained problem into an unconstrained one if such $\lambda^\star$ exists

Lagrangian function

the *Lagrangian function* for problem (8.1) is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i h_i(\boldsymbol{x})$$

- $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ is a p-vector
- the entries of λ_i are called the *Lagrange multipliers*
- the *gradient of Lagrangian* is

$$\nabla L(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) \\ \nabla_\lambda L(\boldsymbol{x}, \boldsymbol{\lambda}) \end{bmatrix}$$

  where

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \nabla h_i(\boldsymbol{x})$$

$$\nabla_\lambda L(\boldsymbol{x}, \boldsymbol{\lambda}) = h(\boldsymbol{x})$$

Method of Lagrange multipliers

Regular point: a feasible point x is a *regular point* if the vectors

$$\nabla h_1(\boldsymbol{x}),\ \nabla h_2(\boldsymbol{x}),\ \ldots,\ \nabla h_p(\boldsymbol{x})$$

are linearly independent

Lagrange theorem: if x^o is a regular point and a local minimizer of the constrained problem (8.1), then there exists a vector $\boldsymbol{\lambda}^o$ such that

$$\nabla_x L(\boldsymbol{x}^o, \boldsymbol{\lambda}^o) = \nabla f(\boldsymbol{x}^o) + \sum_{i=1}^{p} \lambda_i^o \nabla h_i(\boldsymbol{x}^o) = \mathbf{0} \tag{8.2a}$$

$$h(\boldsymbol{x}^o) = \mathbf{0} \tag{8.2b}$$

- there can be *stationary points* (*critical points*), $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\lambda}})$, that satisfy, but $\hat{\boldsymbol{x}}$ is not a local minimizer
- the above method is known as the *method of Lagrange multipliers*

Example 8.2

find the stationary points of the optimization problem:

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1^2 + 2x_2^2 = 1 \end{array}$$

- the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = x_1^2 + x_2^2 + \lambda(x_1^2 + 2x_2^2 - 1)$$

  the necessary optimality conditions are

$$\nabla_x L(\boldsymbol{x}, \lambda) = \begin{bmatrix} 2x_1 + 2x_1\lambda \\ 2x_2 + 4x_2\lambda \end{bmatrix} = \mathbf{0}$$

$$\nabla_\lambda L(\boldsymbol{x}, \lambda) = x_1^2 + 2x_2^2 - 1 = 0$$

solving, we get the stationary points

$$\boldsymbol{x} = (0, \pm\tfrac{1}{\sqrt{2}}), \;\; \lambda = -1/2$$

or

$$\boldsymbol{x} = (\pm 1, 0), \;\; \lambda = -1$$

- all feasible points are regular since $\nabla h(\boldsymbol{x}) = (2x_1, 4x_2)$ is linearly independent for all feasible points; thus, any minimizer to the above problem must satisfy the optimality conditions

- checking the value of the objective, we see that it is smallest at

$$\boldsymbol{x}^{(1)} = (0, \tfrac{1}{\sqrt{2}}) \quad \text{and} \quad \boldsymbol{x}^{(2)} = (0, -\tfrac{1}{\sqrt{2}})$$

- therefore, the points $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are candidate minimizers

Example 8.3

consider the problem of finding the maximum box volume with fixed area $c = 2$:

$$\begin{array}{ll} \text{maximize} & x_1x_2x_3 \\ \text{subject to} & x_1x_2 + x_2x_3 + x_1x_3 = \dfrac{c}{2} \end{array}$$

here, $\boldsymbol{x} = (x_1, x_2, x_3)$ represent the box dimensions

- the gradient of the constraint function $h(\boldsymbol{x}) = x_1x_2 + x_2x_3 + x_1x_3 - 1$ is

$$\nabla h(\boldsymbol{x}) = (x_2 + x_3, x_1 + x_3, x_1 + x_2)$$

  since $\nabla h(\boldsymbol{x}) \neq \mathbf{0}$ for all feasible $\boldsymbol{x}$, all feasible points are regular, and thus, a local solution must satisfy the Lagrange conditions

- the Lagrangian of the equivalent minimization problem is

$$L(\boldsymbol{x}, \lambda) = -x_1x_2x_3 + \lambda(x_1x_2 + x_2x_3 + x_1x_3 - 1)$$

- the necessary optimality conditions are

$$\nabla_x L(\boldsymbol{x}, \lambda) = \begin{bmatrix} -x_2x_3 + \lambda(x_2 + x_3) \\ -x_1x_3 + \lambda(x_1 + x_3) \\ -x_1x_2 + \lambda(x_1 + x_2) \end{bmatrix} = \mathbf{0}$$

$$\nabla_\lambda L(\boldsymbol{x}, \lambda) = x_1x_2 + x_2x_3 + x_1x_3 - 1 = 0$$

  if either one of x_1, x_2, x_3, λ is zero, then the constraint are not satisfied; hence, x_1, x_2, x_3, λ are all nonzero

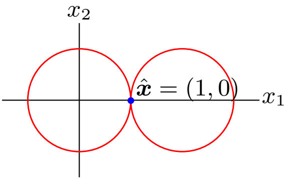
- solving for the above equations, we get $\lambda = \pm\sqrt{3}/6$ and

$$x_1 = x_2 = x_3 = \pm\frac{1}{\sqrt{3}}$$

  since the point $\hat{\boldsymbol{x}} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ has larger objective, it is a local maximizer candidate

Example 8.4

$$\begin{array}{ll} \text{minimize} & x_2 \\ \text{subject to} & x_1^2 + x_2^2 = 1, \\ & (x_1 - 2)^2 + x_2^2 = 1 \end{array}$$



one feasible point $\hat{\boldsymbol{x}} = (1, 0)$, thus optimal

- $(1, 0)$ is not a regular point since $\nabla h_1(\hat{\boldsymbol{x}}) = (2, 0)$ and $\nabla h_2(\hat{\boldsymbol{x}}) = (-2, 0)$ are linearly dependent
- the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = x_2 + \lambda_1(x_1^2 + x_2^2 - 1) + \lambda_2\big((x_1 - 2)^2 + x_2^2 - 1\big)$$

the first necessary condition

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} 2x_1\lambda_1 + 2(x_1 - 2)\lambda_2 \\ 1 + 2x_2(\lambda_1 + \lambda_2) \end{bmatrix} = \mathbf{0}$$

cannot be satisfied at $\hat{\boldsymbol{x}} = (1, 0)$

Second-order conditions: motivation

Lagrange conditions provides necessary conditions and it is still unclear how to check if a stationary point is a local minimizer or not

if the points $\boldsymbol{x}^o, \boldsymbol{\lambda}^o$ satisfy the Lagrange conditions, then, $\boldsymbol{x}^o$ is a stationary point of the unconstrained problem

$$\text{minimize } L(\boldsymbol{x}, \boldsymbol{\lambda}^o)$$

where

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i h_i(\boldsymbol{x})$$

- apply second-order optimality condition for unconstrained problem, that is, we check the definiteness of the Lagrangian Hessain

$$\nabla_x^2 L(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla^2 f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \nabla^2 h_j(\boldsymbol{x})$$

- however, we only need to check the Lagrangian Hessian for feasible directions

Approximate feasible directions

- using Taylor approximation, we can approximate $h_i : \mathbb{R}^n \to \mathbb{R}$ around $\boldsymbol{x}$ by

$$h_i(\boldsymbol{x} + \Delta \boldsymbol{x}) \approx h_i(\boldsymbol{x}) + \nabla h_i(\boldsymbol{x})^T \Delta \boldsymbol{x}$$

  where $\Delta \boldsymbol{x}$ is close to $\boldsymbol{x}$

- if $\boldsymbol{x}$ is feasible ($h_i(\boldsymbol{x}) = 0$), then $\Delta \boldsymbol{x}$ is approximately a feasible direction for $h_i(\boldsymbol{x}) = 0$ if

$$0 = h_i(\boldsymbol{x} + \Delta \boldsymbol{x}) \approx \nabla h_i(\boldsymbol{x})^T \Delta \boldsymbol{x}$$

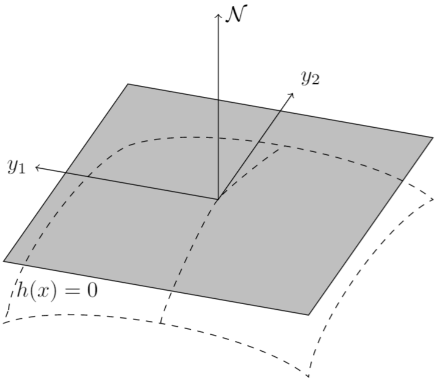
- hence, the *set of approximate feasible directions* is

$$\begin{aligned} \mathcal{T}(\boldsymbol{x}) &= \{\boldsymbol{y} \mid \nabla h_i(\boldsymbol{x})^T \boldsymbol{y} = 0,\ i = 1, \ldots, p\} \\ &= \{\boldsymbol{y} \mid Dh(\boldsymbol{x})\boldsymbol{y} = \boldsymbol{0}\} \end{aligned} \tag{8.3}$$

Tangent space

if $\boldsymbol{x}$ is a regular point then the set of feasible directions $\mathcal{T}(x)$ is a **tangent space** to the surface:

$$\mathcal{S} = \{\boldsymbol{x} \in \mathbb{R}^n \mid h(\boldsymbol{x}) = \mathbf{0}\}$$



Example 8.5

consider the x_3-axis in $\mathbb{R}^3$ constraints:

$$\mathcal{S} = \{\boldsymbol{x} \in \mathbb{R}^3 \mid h_1(\boldsymbol{x}) = x_1 = 0, \quad h_2(\boldsymbol{x}) = x_1 - x_2 = 0\}$$

- we have

$$Dh(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x})^T \\ \nabla h_2(\boldsymbol{x})^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

  the approximate feasible directions, $\boldsymbol{y}$, satisfy

$$Dh(\boldsymbol{x})\boldsymbol{y} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \mathbf{0}$$

  the above holds for $\boldsymbol{y} = (0, 0, \alpha)$ where $\alpha \in \mathbb{R}$; thus, the tangent space is

$$\mathcal{T}(\boldsymbol{x}^o) = \{(0, 0, \alpha) \mid \alpha \in \mathbb{R}\} = \text{the } x_3 \text{ axis in } \mathbb{R}^3$$

Second order conditions: equality constrained case

Necessary conditions: if $\boldsymbol{x}^o$ is a regular point and a local minimizer of problem (8.1), then, there exists a point $\boldsymbol{\lambda}^o$ such that

- $\nabla f(\boldsymbol{x}^o) + \sum_{i=1}^m \nabla h_i(\boldsymbol{x}^o)\lambda_i^o = \mathbf{0}$
- for all $\boldsymbol{y} \in \mathcal{T}(\boldsymbol{x}^o) = \{\boldsymbol{y} \mid Dh(\boldsymbol{x}^o)\boldsymbol{y} = \mathbf{0}\}$, we have

$$\boldsymbol{y}^T \nabla_x^2 L(\boldsymbol{x}^o, \boldsymbol{\lambda}^o)\boldsymbol{y} \geq 0$$

Sufficient conditions: if there exists points $\boldsymbol{x}^o$ and $\boldsymbol{\lambda}^o$ such that

- $\nabla f(\boldsymbol{x}^o) + \sum_{i=1}^m \nabla h_i(\boldsymbol{x}^o)\lambda_i^o = \mathbf{0}$, $h(\boldsymbol{x}^o) = \mathbf{0}$
- for all $\boldsymbol{y} \in \mathcal{T}(\boldsymbol{x}^o) = \{\boldsymbol{y} \mid Dh(\boldsymbol{x}^o)\boldsymbol{y} = \mathbf{0}\}$, $\boldsymbol{y} \neq \mathbf{0}$, we have

$$\boldsymbol{y}^T \nabla_x^2 L(\boldsymbol{x}^o, \boldsymbol{\lambda}^o)\boldsymbol{y} > 0,$$

then, $\boldsymbol{x}^o$ is a strict local minimizer of problem (8.1)

Example 8.6

$$\begin{array}{ll} \text{minimize} & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 3 \end{array}$$

find the stationary points and determine whether they are local minimizers

- the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = x_1x_2 + x_2x_3 + x_1x_3 + \lambda(x_1 + x_2 + x_3 - 3)$$

  the first-order necessary conditions are

$$\nabla_x L(\boldsymbol{x}, \lambda) = \begin{bmatrix} x_2 + x_3 + \lambda \\ x_1 + x_3 + \lambda \\ x_1 + x_2 + \lambda \end{bmatrix} = \mathbf{0}$$

$$x_1 + x_2 + x_3 = 3$$

  and the solution is $x_1 = x_2 = x_3 = 1, \lambda = -2$

- to check whether the point $\hat{\boldsymbol{x}} = (1, 1, 1)$ is a local minimizer, we look at the second-order condition
- note that $\nabla h(\boldsymbol{x}) = (1, 1, 1)$ and the Hessian

$$\nabla_x^2 L(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

is an indefinite matrix; however, on the tangent space

$$\mathcal{T} = \{\boldsymbol{y} \mid \nabla h(\hat{\boldsymbol{x}})^T \boldsymbol{y} = 0\} = \{\boldsymbol{y} \mid y_1 + y_2 + y_3 = 0\}$$

we have

$$\begin{aligned} \boldsymbol{y}^T \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \boldsymbol{y} &= y_1(y_2 + y_3) + y_2(y_1 + y_3) + y_3(y_1 + y_2) \\ &= -(y_1^2 + y_2^2 + y_3^2) < 0, \end{aligned}$$

which is negative definite; thus, the solution $\hat{\boldsymbol{x}} = (1, 1, 1)$ is not a local minimizer (it is a local maximizer)

Quadratic objective and constraint

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^T Q \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T P \boldsymbol{x} = 1 \end{array}$$

where $Q = Q^T$ and $P = P^T > 0$

- the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = \boldsymbol{x}^T Q \boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T P \boldsymbol{x})$$

- the Lagrange conditions are

$$\nabla_x L(\boldsymbol{x}, \lambda) = 2Q\boldsymbol{x} - 2\lambda P \boldsymbol{x} = \mathbf{0}$$
$$\nabla_\lambda L(\boldsymbol{x}, \lambda) = 1 - \boldsymbol{x}^T P \boldsymbol{x} = 0$$

- from the first equation, we have

$$P^{-1}Q\boldsymbol{x} = \lambda\boldsymbol{x}$$

hence, a solution $\hat{\boldsymbol{x}}$ and $\hat{\lambda}$ if they exists, are eigenvectors and eigenvalues of $P^{-1}Q$

- multiplying the equation $P^{-1}Q\boldsymbol{x} = \lambda\boldsymbol{x}$ on the left by $\boldsymbol{x}^T P$ and using $\boldsymbol{x}^T P\boldsymbol{x} = 1$, we get

$$\lambda = \boldsymbol{x}^T Q\boldsymbol{x} = f(\boldsymbol{x})$$

- hence, $f(\boldsymbol{x}) = \boldsymbol{x}^T Q\boldsymbol{x} = \lambda$ is minimized when λ is the smallest eigenvalue of $P^{-1}Q$ and $\boldsymbol{x}$ is the corresponding eigenvector, which is a minimizer

Example 8.7

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^T Q \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T P \boldsymbol{x} = 1 \end{array}$$

where

$$Q = \begin{bmatrix} -4 & 0 \\ 0 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

- the minimum eigenvalue of

$$P^{-1}Q = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

is $\hat{\lambda} = -2$; substituting, $\lambda = -2$ in the Lagrange conditions, we have

$$\nabla_x L(\boldsymbol{x}, -2) = 2Q\boldsymbol{x} - 2\lambda P \boldsymbol{x} = \begin{bmatrix} 0 \\ 2x_2 \end{bmatrix} = \mathbf{0}$$

$$\nabla_\lambda L(\boldsymbol{x}, -2) = 1 - 2x_1^2 - x_2^2 = 0$$

- solving, we get the solutions $\hat{\boldsymbol{x}}_1 = (1/\sqrt{2}, 0)$ or $\hat{\boldsymbol{x}}_2 = (-1/\sqrt{2}, 0)$

- to verify that these points are strict local minimizers, we find the Hessian of the Lagrangian (for first $\hat{\boldsymbol{x}}_1$, the other follow similar steps)

$$\nabla_x^2 L(\boldsymbol{x}, \hat{\lambda}) = 2Q - 2\hat{\lambda}P = 2(Q + 2P) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

- since $h(\boldsymbol{x}) = 1 - \boldsymbol{x}^T P \boldsymbol{x} = 0$, we have $\nabla h(\boldsymbol{x}) = -2P\boldsymbol{x}$ and the tangent space is

$$\mathcal{T}(\hat{\boldsymbol{x}}) = \{\boldsymbol{y} \mid 2\hat{\boldsymbol{x}}^T P \boldsymbol{y} = \boldsymbol{0}\} = \{\boldsymbol{y} \mid [\sqrt{2}, 0]\boldsymbol{y} = \boldsymbol{0}\} = \{(0, a) \mid a \in \mathbb{R}\}$$

- for every $\boldsymbol{y} \in \mathcal{T}$, $\boldsymbol{y} \neq \boldsymbol{0}$, we have

$$\boldsymbol{y}^T \nabla_x^2 L(\hat{\boldsymbol{x}}, \hat{\lambda}) \boldsymbol{y} = 2a^2 > 0$$

  we conclude that the point $\hat{\boldsymbol{x}} = (\frac{1}{\sqrt{2}}, 0)$ is a local minimizer

Outline

- equality constrained problems
- **inequality constrained problems**
- quadratic problems with linear constraints
- projected gradient descent
- penalty method

Inequality constrained problems

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, p \end{array} \tag{8.4}$$

- $f : \mathbb{R}^n \to \mathbb{R}$
- $g_i : \mathbb{R}^n \to \mathbb{R}$
- $h_j : \mathbb{R}^n \to \mathbb{R}$
- $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x}))$
- $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_p(\boldsymbol{x}))$
- $\hat{\boldsymbol{x}}$ is *a feasible point* if it satisfies the constraints ($g(\hat{\boldsymbol{x}}) \leq \mathbf{0},\ h(\hat{\boldsymbol{x}}) = \mathbf{0}$)

Motivation of optimality conditions

if $\boldsymbol{x}^o$ is a local minimizer of (8.4), then it is a local minimizer of the problem:

$$\begin{array}{ll}\text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) = \boldsymbol{0},\ i \in \mathcal{I}(\boldsymbol{x}^o),\ h(\boldsymbol{x}) = \boldsymbol{0}\end{array}$$

- using Lagrange conditions (8.2) on the above problem, we have

$$\nabla f(\boldsymbol{x}^o) + \sum_{i \in \mathcal{I}(\boldsymbol{x}^o)} \mu_i^o \nabla g_i(\boldsymbol{x}^o) + \sum_{j=1}^{p} \lambda_j^o \nabla h_j(\boldsymbol{x}^o) = \boldsymbol{0}$$

- in terms of the original problem, we can write the above condition as

$$\nabla f(\boldsymbol{x}^o) + \sum_{i=1}^{m} \mu_i^o \nabla g_i(\boldsymbol{x}^o) + \sum_{j=1}^{p} \lambda_j^o \nabla h_j(\boldsymbol{x}^o) = \boldsymbol{0}$$

$$\mu_i = 0 \text{ for } i \notin \mathcal{I}(\boldsymbol{x}^o) \Rightarrow g_i(\boldsymbol{x}^o)^T \mu_i^o = 0$$

  it can be shown that $\mu_i \geq 0$ for $i \in \mathcal{I}(\boldsymbol{x}^o)$

Lagrangian

the *Lagrangian* associated with problem (8.4) is

$$L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{p} \lambda_j h_j(\boldsymbol{x})$$

- $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$
- both $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are often called Lagrange multipliers vectors
- the gradient of the Lagrangian with respect to $\boldsymbol{x}$ is

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \nabla f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}) + \sum_{j=1}^{p} \lambda_j \nabla h_j(\boldsymbol{x})$$

Regular point

Active inequalities

- an inequality constraint $g_i(\boldsymbol{x}) \leq 0$ is *active* at $\hat{\boldsymbol{x}}$ if $g_i(\hat{\boldsymbol{x}}) = 0$
- it is *inactive* at $\hat{\boldsymbol{x}}$ if $g_i(\hat{\boldsymbol{x}}) < 0$
- we let $\mathcal{I}(\hat{\boldsymbol{x}})$ denote the set of indices i for the active constraints at $\hat{\boldsymbol{x}}$:

$$\mathcal{I}(\hat{\boldsymbol{x}}) = \{i \mid g_i(\hat{\boldsymbol{x}}) = 0\}$$

Regular point: a feasible point $\hat{\boldsymbol{x}}$ is a *regular point* if the vectors

$$\nabla g_i(\hat{\boldsymbol{x}}),\ \nabla h_j(\hat{\boldsymbol{x}}), \quad i \in \mathcal{I}(\hat{\boldsymbol{x}}),\ j = 1, \ldots, p$$

are linearly independent

Karush-Kuhn-Tucker (KKT) conditions

if $\boldsymbol{x}^o$ is a regular point and a local minimizer for problem (8.4), then there exists $\boldsymbol{\mu}^o \in \mathbb{R}^m$ and $\boldsymbol{\lambda}^o \in \mathbb{R}^p$ such that:

$$\nabla_x L(\boldsymbol{x}^o, \boldsymbol{\mu}^o, \boldsymbol{\lambda}^o) = \mathbf{0} \tag{8.5a}$$

$$g_i(\boldsymbol{x}^o) \leq 0, \quad i = 1, \ldots, m \tag{8.5b}$$

$$h_j(\boldsymbol{x}^o) = 0, \quad j = 1, \ldots, p \tag{8.5c}$$

$$\mu_i^o \geq 0, \quad i = 1, \ldots, m \tag{8.5d}$$

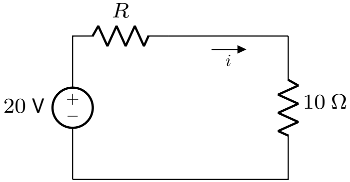
$$\mu_i^o g_i(\boldsymbol{x}^o) = 0, \quad i = 1, \ldots, m \tag{8.5e}$$

the vectors $\boldsymbol{\lambda}^o$ and $\boldsymbol{\mu}^o$ are called the *Lagrange multiplier* and *KKT multiplier* vectors (or just Lagrange multiplier vectors)

Complementary slackness: the last KKT condition $\mu_i^o g_i(\boldsymbol{x}^o) = 0$ is called the complementary slackness; it implies that

- $g_i(\boldsymbol{x}^o) < 0 \Rightarrow \mu_i^o = 0$
- $\mu_i^o > 0 \Rightarrow g_i(\boldsymbol{x}^o) = 0$

Example 8.8



let us determine the value of the resistor $R \geq 0$ such that the power absorbed by this resistor is maximized

the power absorbed R is $p = i^2 R$ where $i = 20/(10+R)$; hence, the problem can formulated as

$$\begin{array}{ll} \text{minimize} & -\frac{400x}{(10+x)^2} \\ \text{subject to} & -x \leq 0 \end{array}$$

the variable x represents the resistor R

the Lagrangian is

$$L(x, \mu) = -\frac{400x}{(10+x)^2} - \mu x$$

the derivative of the objective function is

$$-\frac{400(10+x)^2 - 800x(10+x)}{(10+x)^4} = -\frac{400(10-x)}{(10+x)^3}$$

KKT conditions:

$$\begin{aligned} -\frac{400(10-x)}{(10+x)^3} - \mu = 0 \\ \mu \geq 0 \\ \mu x = 0 \\ -x \leq 0 \end{aligned}$$

- if $\mu > 0$, then $x = 0$, and the first equation does not hold
- let $\mu = 0$; then we get $x = 10$, which satisfies all conditions
- hence, the point $x = 10$ is a stationary point and a local minimizer candidate

Example 8.9

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 + x_1 x_2 - 3x_1 \\ \text{subject to} & x_1 \geq 0, x_2 \geq 0 \end{array}$$

- the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = x_1^2 + x_2^2 + x_1 x_2 - 3x_1 - \mu_1 x_1 - \mu_2 x_2$$

- note that $g(\boldsymbol{x}) = (-x_1, -x_2)$ and the KKT conditions are

$$\begin{aligned} \nabla_x L(\boldsymbol{x}, \boldsymbol{\mu}) = \begin{bmatrix} 2x_1 + x_2 - 3 - \mu_1 \\ x_1 + 2x_2 - \mu_2 \end{bmatrix} &= \mathbf{0} \\ \boldsymbol{\mu} &\geq \mathbf{0} \\ -\boldsymbol{x} &\leq \mathbf{0} \\ \mu_1 x_1 &= 0 \\ \mu_2 x_2 &= 0 \end{aligned}$$

- to find a solution, suppose that $\mu_1 = 0$ and $x_2 = 0$; then, solving the above with these values, we have

$$\boldsymbol{x} = \begin{bmatrix} \frac{3}{2} \\ 0 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}$$

  which satisfy the KKT conditions

- if we try $\mu_2 = 0$ and $x_1 = 0$, we get $x_2 = 0$, $\mu_1 = -3$, which violates the condition $\boldsymbol{\mu} \geq \mathbf{0}$

- similarly, the other combinations $x_1 = x_2 = 0$ and $\mu_1 = \mu_2 = 0$ violates the KKT condition

Necessary conditions: inequality constrained case

Tangent space

$$\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{y} \mid Dh(\boldsymbol{x})\boldsymbol{y} = \mathbf{0},\ \nabla g_i(\boldsymbol{x})^T\boldsymbol{y} = 0,\ i \in \mathcal{I}(\boldsymbol{x})\}$$

- $\mathcal{I}(\boldsymbol{x}) = \{i \mid g_i(\boldsymbol{x}) = 0\}$ is the set with active constraints indices
- tangent space is the set of feasible directions with active constraints

Necessary conditions: suppose $\boldsymbol{x}^o$ is a regular point and a local minimizer of problem (8.4), then, there exists $\boldsymbol{\mu}^o, \boldsymbol{\lambda}^o$ such that:

- the KKT conditions (8.5) hold; and
- for all $\boldsymbol{y} \in \mathcal{T}(\boldsymbol{x}^o)$, we have

$$\boldsymbol{y}^T\nabla_x^2 L(\boldsymbol{x}^o, \boldsymbol{\mu}^o, \boldsymbol{\lambda}^o)\boldsymbol{y} \geq 0$$

Sufficient conditions: inequality constrained case

Critical tangent space: for any points x, $\boldsymbol{\mu}$, and $\boldsymbol{\lambda}$ satisfying the KKT conditions (8.5), we define the *critical tangent space* as:

$$\bar{\mathcal{T}}(\boldsymbol{x}) = \{\boldsymbol{y} \mid Dh(\boldsymbol{x})\boldsymbol{y} = \boldsymbol{0},\ \nabla g_i(\boldsymbol{x})^T \boldsymbol{y} = 0,\ i \in \bar{\mathcal{I}}(\boldsymbol{x})\}$$

where $\bar{\mathcal{I}}(\boldsymbol{x}) = \{i \mid g_i(\boldsymbol{x}) = 0, \mu_i > 0\}$

Sufficient conditions: suppose that there exists points $\boldsymbol{x}^o$, $\boldsymbol{\mu}^o$, and $\boldsymbol{\lambda}^o$ such that the KKT conditions (8.5) hold; if for all $\boldsymbol{y} \in \bar{\mathcal{T}}(\boldsymbol{x}^o)$, $\boldsymbol{y} \neq \boldsymbol{0}$, we have

$$\boldsymbol{y}^T \nabla_x^2 L(\boldsymbol{x}^o, \boldsymbol{\lambda}^o, \boldsymbol{\mu}^o)\boldsymbol{y} > 0,$$

then, $\boldsymbol{x}^o$ is a strict local minimizer of (8.4)

Example 8.10

$$\begin{array}{ll} \text{minimize} & x_1 x_2 \\ \text{subject to} & x_1 + x_2 \geq 2, \quad x_1 - x_2 \leq 0 \end{array}$$

- the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = x_1 x_2 + \mu_1(2 - x_1 - x_2) + \mu_2(x_1 - x_2)$$

- we have $g_1(\boldsymbol{x}) = 2 - x_1 - x_2$ and $g_2(\boldsymbol{x}) = x_1 - x_2$ and the KKT conditions are

$$\begin{aligned} \nabla_x L(\boldsymbol{x}, \boldsymbol{\mu}) = \begin{bmatrix} x_2 - \mu_1 + \mu_2 \\ x_1 - \mu_1 - \mu_2 \end{bmatrix} &= \mathbf{0} \\ 2 - x_1 - x_2 &\leq 0 \\ x_1 - x_2 &\leq 0 \\ \mu_1, \mu_2 &\geq 0 \\ \mu_1(2 - x_1 - x_2) &= 0 \\ \mu_2(x_1 - x_2) &= 0 \end{aligned}$$

- it can be verified that $\mu_1 \neq 0$ and $\mu_2 = 0$; solving with $\mu_2 = 0$, we arrive at one solution: $\hat{x}_1 = \hat{x}_2 = 1, \mu_1 = 1, \mu_2 = 0$

- at this solution, the constraints are active, and

$$\nabla g_1(\hat{\boldsymbol{x}}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla g_2(\hat{\boldsymbol{x}}) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \nabla_x^2 L(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\mu}}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

  the vectors $\nabla g_1(\hat{\boldsymbol{x}}), \nabla g_2(\hat{\boldsymbol{x}})$ are linearly independent, hence $\hat{\boldsymbol{x}}$ is regular

- since both constraints are active, the tangent space is

$$\mathcal{T} = \{\boldsymbol{y} \mid \nabla g_1(\hat{\boldsymbol{x}})^T \boldsymbol{y} = 0,\ \nabla g_2(\hat{\boldsymbol{x}})^T \boldsymbol{y} = 0\} = \{\boldsymbol{0}\}$$

  therefore, $\boldsymbol{y}^T \nabla_x^2 L(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\mu}}) \boldsymbol{y} = 0$ for $\boldsymbol{y} \in \mathcal{T}$ and the point $\hat{\boldsymbol{x}}$ is a candidate local minimizer

- we now check the sufficient conditions; since $\mu_2 = 0$, the critical tangent space is

$$\begin{aligned}\bar{\mathcal{T}} &= \{\boldsymbol{y} \mid \nabla g_1(\hat{\boldsymbol{x}})^T \boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} \mid -y_1 - y_2 = 0\} \\ &= \{\boldsymbol{y} \mid y_1 = -y_2\}\end{aligned}$$

- for $\boldsymbol{y} \in \bar{\mathcal{T}}$, $\boldsymbol{y} \neq \mathbf{0}$, we have

$$\boldsymbol{y}^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{y} = 2y_1 y_2 = -2y_2^2 < 0$$

  this means that the sufficient condition does not hold

- hence, $\hat{\boldsymbol{x}}$ is not a local minimizer (it is also not a local maximizer)

Example 8.11

$$
\begin{array}{ll}
\text{minimize} & (x_1 - 1)^2 + x_2 - 2 \\
\text{subject to} & x_2 = x_1 + 1, \quad x_1 + x_2 \leq 2
\end{array}
$$

- we have $h(\boldsymbol{x}) = x_2 - x_1 - 1$ and $g(\boldsymbol{x}) = x_1 + x_2 - 2$ and

$$
\nabla h(\boldsymbol{x}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \nabla g(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix},
$$

  are linearly independent; hence, all feasible points are regular and a local solution must satisfy the KKT conditions

- the Lagrangian is

$$
L(\boldsymbol{x}, \mu, \lambda) = (x_1 - 1)^2 + x_2 - 2 + \mu(x_1 + x_2 - 2) + \lambda(x_2 - x_1 - 1)
$$

KKT conditions:

$$
\begin{aligned}
\begin{bmatrix} 2x_1 - 2 + \mu - \lambda \\ 1 + \mu + \lambda \end{bmatrix} &= \mathbf{0} \\
\mu(x_1 + x_2 - 2) &= 0 \\
\mu &\geq 0 \\
x_2 - x_1 - 1 &= 0 \\
x_1 + x_2 - 2 &\leq 0
\end{aligned}
$$

- for $\mu > 0$, we will get an invalid solution; solving with $\mu = 0$, we arrive at the solution
$$x_1 = \tfrac{1}{2}, \quad x_2 = \tfrac{3}{2}, \quad \lambda = -1$$
- the point $\hat{\boldsymbol{x}} = (\frac{1}{2}, \frac{3}{2})$ is a local minimizer candidate

- the Hessian of the Lagrangian is

$$\nabla_x^2 L(\boldsymbol{x}, \mu, \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

  for all $\boldsymbol{x}$ (positive semi-definite)

- since $\mu = 0$, the critical tangent space is:

$$\begin{aligned} \bar{\mathcal{T}} = \{\boldsymbol{y} \mid \nabla h(\hat{\boldsymbol{x}})^T \boldsymbol{y} = 0\} &= \{\boldsymbol{y} \mid -y_1 + y_2 = 0\} \\ &= \{\boldsymbol{y} = (a, a) \mid a \in \mathbb{R}\} \end{aligned}$$

- for $\boldsymbol{y} \in \bar{\mathcal{T}}$, we have

$$\boldsymbol{y}^T \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{y} = 2a^2 > 0,$$

  which is positive-definite; therefore, the point $\hat{\boldsymbol{x}}$ is a local minimizer

Outline

- equality constrained problems
- inequality constrained problems
- **quadratic problems with linear constraints**
- projected gradient descent
- penalty method

Quadratic program with linear constraints

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x} \\ \text{subject to} & C\boldsymbol{x} = \boldsymbol{d} \end{array}$$

- Q is an $n \times n$ symmetric matrix; $\boldsymbol{r}$ is an n-vector
- C is a $p \times n$ matrix; $\boldsymbol{d}$ is a p-vector

the Lagrangian for this problem is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (C\boldsymbol{x} - \boldsymbol{d})$$

Solution

a solution (if it exists) must satisfy the following Lagrange optimality conditions:

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = Q\boldsymbol{x} + \boldsymbol{r} + C^T\boldsymbol{\lambda} = \boldsymbol{0} \tag{8.6a}$$

$$C\boldsymbol{x} - \boldsymbol{d} = \boldsymbol{0} \tag{8.6b}$$

the above can be written as the system of linear equations:

$$\begin{bmatrix} Q & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{r} \\ \boldsymbol{d} \end{bmatrix} \tag{8.7}$$

- the solution of the above can be a minimizer, maximizer, or a saddle point
- if Q is positive semidefinite, then any solution of the above is a global minimizer

Closed-form solution: assume Q is invertible and C has linearly independent rows

- multiply the first equation in (8.6) by Q^{-1} on the left

$$\boldsymbol{x} = -Q^{-1}(\boldsymbol{r} + C^T\boldsymbol{\lambda})$$

- substituting into the second equation, we get

$$-CQ^{-1}(\boldsymbol{r} + C^T\boldsymbol{\lambda}) = \boldsymbol{d} \iff \big(CQ^{-1}C^T\big)\boldsymbol{\lambda} = -(\boldsymbol{d} + CQ^{-1}\boldsymbol{r})$$

  hence

$$\boldsymbol{\lambda} = -\big(CQ^{-1}C^T\big)^{-1}(\boldsymbol{d} + AQ^{-1}\boldsymbol{r})$$

- putting it all together, we get

$$\boldsymbol{x} = Q^{-1}C^T\big(CQ^{-1}C^T\big)^{-1}(CQ^{-1}\boldsymbol{r} + \boldsymbol{d}) - Q^{-1}\boldsymbol{r}$$

Example 8.12

consider the discrete-time linear system

$$s_k = 2s_{k-1} + u_k, \quad k \geq 1,$$

with $s_0 = 1$; suppose that we want to find the values of the inputs u_1 and u_2 that minimizes

$$\frac{1}{2}u_1^2 + \frac{1}{3}u_2^2 + s_2^2$$

- we can formulate this problem as a quadratic program with variables u_1, u_2 and s_2
- the state at time 2 can be found recursively as:

$$s_2 = 2s_1 + u_2 = 2(2s_0 + u_1) + u_2 = 2(2 + u_1) + u_2$$

hence,

$$2u_1 + u_2 - s_2 = -4$$

therefore, the problem can be formulated as:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}u_1^2 + \frac{1}{3}u_2^2 + s_2^2 \\ \text{subject to} & 2u_1 + u_2 - s_2 = -4 \end{array}$$

letting $\boldsymbol{x} = (u_1, u_2, s_2)$, we can write the problem as:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} \\ \text{subject to} & C\boldsymbol{x} = d \end{array}$$

where

$$Q = \operatorname{diag}(1,\ 2/3,\ 2), \quad C = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix}, \quad d = -4$$

this is a quadratic problem with linear constraints; since Q is invertible and C is a nonzero row vector, the solution is

$$\boldsymbol{x} = (u_1, u_2, s_2) = Q^{-1}C^T(CQ^{-1}C^T)^{-1}d = \left(-\tfrac{4}{3}, -1, \tfrac{1}{3}\right)$$

Constrained least squares

$$\begin{array}{ll} \text{minimize} & \|A\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{subject to} & C\boldsymbol{x} = \boldsymbol{d} \end{array}$$

where A is an $m \times n$ matrix, C is a $p \times n$ matrix, $\boldsymbol{b}$ is an m-vector, and $\boldsymbol{d}$ is a p-vector

- the objective is $\|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \boldsymbol{x}^T(A^TA)\boldsymbol{x} - 2(A^T\boldsymbol{b})^T\boldsymbol{x} + \|\boldsymbol{b}\|^2$
- quadratic objective with $Q = 2A^TA$, $\boldsymbol{r} = -2A^T\boldsymbol{b}$
- hence, the optimality condition is

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 2A^T\boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}$$

- $Q = 2A^TA \geq 0$; so any solution of the above is a global minimizer

Outline

- equality constrained problems
- inequality constrained problems
- quadratic problems with linear constraints
- **projected gradient descent**
- penalty method

Projection

Constrained optimization

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \end{array}$$

- $\boldsymbol{x} \in \mathbb{R}^n$ is variable; $f : \mathbb{R}^n \to \mathbb{R}$
- $\mathcal{X}$ is the constraint set

Projection: the *projection of* $\boldsymbol{x} \in \mathbb{R}^n$ *onto* $\mathcal{X} \subseteq \mathbb{R}^n$ is

$$\Pi_{\mathcal{X}}[\boldsymbol{x}] = \underset{\boldsymbol{z} \in \mathcal{X}}{\operatorname{argmin}} \|\boldsymbol{z} - \boldsymbol{x}\|$$

- the point $\Pi_{\mathcal{X}}[\boldsymbol{x}]$ is the "closest" point in $\mathcal{X}$ to $\boldsymbol{x}$
- for certain constraints, the projection can be computed in closed form

Examples

- *Box constraint*

$$\mathcal{X} = \{\boldsymbol{x} \mid l_i \leq x_i \leq u_i,\ i = 1, \ldots, n\}$$

given $\boldsymbol{x}$, its projection $\boldsymbol{y} = \Pi_{\mathcal{X}}[\boldsymbol{x}]$ onto $\mathcal{X}$ is

$$y_i = \begin{cases} u_i & \text{if } x_i > u_i \\ x_i & \text{if } l_i \leq x_i \leq u_i \\ l_i & \text{if } x_i < l_i \end{cases}$$

- *Unit ball constraint*

$$\mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|^2 = 1\}$$

the projection is simply the normalization of $\boldsymbol{x}$:

$$\Pi_{\mathcal{X}}[\boldsymbol{x}] = \boldsymbol{x}/\|\boldsymbol{x}\|$$

Gradient descent and projection

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X} \end{array}$$

the gradient descent update has the form:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})$$

- the point $\boldsymbol{x}^{(k+1)}$ is not guaranteed to be in $\mathcal{X}$ even if $\boldsymbol{x}^{(k)}$ is
- to guarantee feasibility, we can modify the update to

$$\boldsymbol{x}^{(k+1)} = \Pi_{\mathcal{X}}\left[\boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})\right]$$

  where $\Pi_{\mathcal{X}}[\boldsymbol{x}]$ denote the projection of $\boldsymbol{x}$ onto $\mathcal{X}$

Projected gradient descent

Algorithm Projected gradient descent

given a starting point $\boldsymbol{x}^{(0)}$ and a solution tolerance $\epsilon > 0$

repeat for $k \geq 1$

1. choose a stepsize α_k
2. update $\boldsymbol{x}^{(k+1)}$:

$$\boldsymbol{x}^{(k+1)} = \Pi_{\mathcal{X}}\left[\boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})\right]$$

if $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\| \leq \epsilon$ stop and $\boldsymbol{x}^{(k+1)}$ is output

$$\Pi_{\mathcal{X}}[\boldsymbol{x}] = \underset{\boldsymbol{z} \in \mathcal{X}}{\operatorname{argmin}} \|\boldsymbol{z} - \boldsymbol{x}\|$$

Examples

- the projected gradient descent update for the problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1 \end{array}$$

  is

$$\boldsymbol{x}^{(k+1)} = \frac{1}{\|(I - \alpha_k Q)\boldsymbol{x}^{(k)}\|}(I - \alpha_k Q)\boldsymbol{x}^{(k)}$$

- the projected gradient descent update for the problem

$$\begin{array}{ll} \text{minimize} & (1/2)\boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \geq \mathbf{0} \end{array}$$

  is

$$\boldsymbol{x}^{(k+1)} = [\boldsymbol{x}^{(k)} - \alpha(Q\boldsymbol{x}^{(k)} + \boldsymbol{r})]_+,$$

  where $[\cdot]_+$ replaces negative entries by zero

Outline

- equality constrained problems
- inequality constrained problems
- quadratic problems with linear constraints
- projected gradient descent
- **penalty method**

Penalized formulation

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, \quad i = 1, \ldots, p \end{array}$$

Penalized formulation

$$\text{minimize} \quad f(\boldsymbol{x}) + \rho P(h(\boldsymbol{x}))$$

- $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \ldots, h_p(\boldsymbol{x}))$
- $P : \mathbb{R}^p \to \mathbb{R}$ is the *penalty function*
- $\rho \in \mathbb{R}$ is the *penalty parameter*
- the role of the term $\rho P(\boldsymbol{x})$ is to penalize constraints violation, *i.e.*, has large values for infeasible points

Penalty function

Penalty function: the penalty function P satisfies the following conditions:

1. P is continuous
2. $P(h(\boldsymbol{x})) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$
3. $P(h(\boldsymbol{x})) = 0$ if and only if $\boldsymbol{x}$ is feasible ($h(\boldsymbol{x}) = \mathbf{0}$)

Quadratic penalty function

$$P(h(\boldsymbol{x})) = \|h(\boldsymbol{x})\|^2 = \sum_{i=1}^{p} \big(h_i(\boldsymbol{x})\big)^2$$

Quadratic penalty formulation

$$\text{minimize} \quad f(\boldsymbol{x}) + \rho\|h(\boldsymbol{x})\|^2$$

- a solution of the above problem might not feasible
- for large ρ we expect to have small values $\big(h_i(\boldsymbol{x})\big)^2$, *i.e.*, an approximate solution to the original problem
- minimizing the penalty problem for an increasing sequence of values of ρ is known as the penalty method

Quadratic penalty method

Algorithm Quadratic penalty method

---

given a starting point $\boldsymbol{x}^{(0)}$, ρ_0, and a solution tolerance $\epsilon > 0$

repeat for $k = 1, 2, \ldots, K$

1. set $\boldsymbol{x}^{(k+1)}$ to be the (approximate) minimizer of

$$\text{minimize} \quad f(\boldsymbol{x}) + \rho_k \|h(\boldsymbol{x})\|^2$$

   using an unconstrained optimization method with initial point $\boldsymbol{x}^{(k)}$

2. update $\rho_{k+1} = 2\rho_k$

- terminate if $\|g^+(\boldsymbol{x})\|^2$ and $\|h(\boldsymbol{x})\|^2$ are small enough
- simple and easy to implement
- but has a major issue: the parameter ρ_k rapidly increases with iterations; when solving penalty problem using gradient descent for example, it can be very slow or simply fail

Inequality constraints

for problems of the form

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \le 0, \quad i = 1, \dots, m \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, p \end{array}$$

we can for example consider the penalized problem:

$$\text{minimize} \quad f(\boldsymbol{x}) + \rho \|h(\boldsymbol{x})\|^2 + \rho \|g^+(\boldsymbol{x})\|^2$$

- $g^+(\boldsymbol{x}) = (g_1^+(\boldsymbol{x}), \dots, g_m^+(\boldsymbol{x}))$ and

$$g_i^+(\boldsymbol{x}) = \max\{0, g_i(\boldsymbol{x})\} = \begin{cases} 0 & \text{if } g_i(\boldsymbol{x}) \le 0 \\ g_i(\boldsymbol{x}) & \text{if } g_i(\boldsymbol{x}) > 0 \end{cases}$$

- there are many other choices of penalty functions; here, we just consider the simple quadratic penalization function

References and further readings

- Edwin KP Chong and Stanislaw H Zak. *An Introduction to Optimization,* John Wiley & Sons, 2013, chapters 20, 21.