8. Constrained optimization

- equality constrained problems
- constrained quadratic problems
- inequality constrained problems
- projected gradient descent

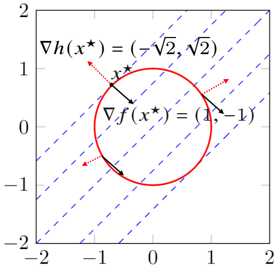
Equality constrained problems

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

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function
- $h_i : \mathbb{R}^n \to \mathbb{R}$ are the equality constraints functions
- we let $h(x) = (h_1(x), \ldots, h_p(x))$
- a point x satisfying $h(x) = 0$ is called a *feasible point*

Example

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



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

$$\nabla f(x^\star) = -\lambda \nabla h(x^\star) \quad \text{where} \quad \lambda = 1/\sqrt{2}$$

Motivation of optimality conditions

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

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

- $\lambda \in \mathbb{R}$ is an adjustable parameter
- assume for some $\lambda^\star$, $x^\star$ minimizes $f(x) + \lambda^\star h(x)$ and satisfies the constraint:

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

- then, we have

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

  hence, $f(x^\star) \leq f(x)$ for all feasible x and $x^\star$ is a solution to the original problem
- we can transform constrained problem into an unconstrained one if such $\lambda^\star$ exists

Lagrangian function

the *Lagrangian function* for problem (8.1) is

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

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

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

  where

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

Optimality conditions

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

$$\nabla h_1(x),\ \nabla h_2(x),\ \ldots,\ \nabla h_p(x)$$

are linearly independent (*i.e.*, $Dh(x)$ has linearly independent rows)

(Lagrange) Optimality conditions: if x° is a regular point and a local minimizer of the constrained problem (8.1), then there exists a vector λ° such that

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

$$h(x^\circ) = 0 \qquad (8.2b)$$

- conditions are necessary but not sufficient
- points that satisfies the above are called *stationary points*
- there can be *stationary points*, $(\hat{x}, \hat{\lambda})$, but $\hat{x}$ is not a local minimizer
- the above method is known as the *method of Lagrange multipliers*

Example

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

- the Lagrangian is

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

- the necessary optimality conditions are

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

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

- solving, we get the stationary points

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

  or

$$x = (\pm 1, 0), \quad \lambda = -1$$

- $\nabla h(x) = (2x_1, 4x_2)$ is linearly independent for all feasible points
- so, all feasible points are regular and any minima satisfies the optimality conditions
- checking the value of the objective, we see that it is smallest at

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

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

Example

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

$$\begin{array}{ll} \text{maximize} & x_1 x_2 x_3 \\ \text{subject to} & x_1 x_2 + x_2 x_3 + x_1 x_3 = \dfrac{c}{2} \end{array}$$

- here, $x = (x_1, x_2, x_3)$ represent the box dimensions
- the gradient of the constraint function $h(x) = x_1 x_2 + x_2 x_3 + x_1 x_3 - 1$ is

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

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

- the Lagrangian of the equivalent minimization problem is

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

- the necessary optimality conditions are

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

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

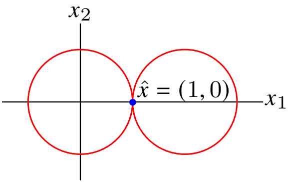
- x_i, λ are nonzero as otherwise, the conditions cannot be met
- solving for the above equations, we get $\lambda = \pm\sqrt{3}/6$ and

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

point $\hat{x} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ has larger objective and is a local maximizer candidate

Example

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



- one feasible point $\hat{x} = (1,0)$, thus optimal
- $(1,0)$ is not regular since $\nabla h_1(\hat{x}) = (2,0)$, $\nabla h_2(\hat{x}) = (-2,0)$ are dependent
- the Lagrangian is

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

- the first necessary condition

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

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

Second-order conditions: motivation

if x°, λ° satisfy the optimality conditions, then, x° is a stationary point of

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

- apply second-order optimality condition for unconstrained problem
- we check the definiteness of the Hessian of the Lagrangian

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

- 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 x by

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

  where Δx is close to x

- if x is feasible ($h_i(x) = 0$), then Δx is approximately a feasible direction if

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

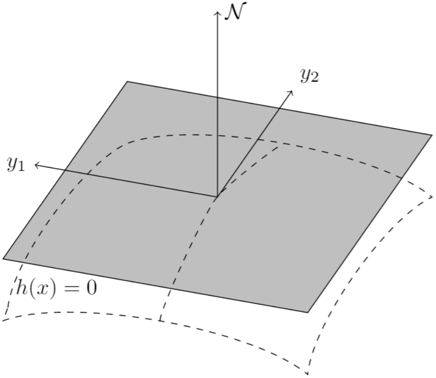
- hence, the *set of approximate feasible directions* is

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

Tangent space

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

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



Example

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

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

- we have

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

- the approximate feasible directions, y, satisfy

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

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

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

Second order optimality conditions

Necessary conditions

if x° is a regular point and a local minimizer, then there exists λ° such that

$$\nabla_x L(x^\circ, \lambda^\circ) = \nabla f(x^\circ) + \sum_{i=1}^{m} \nabla h_i(x^\circ)\lambda_i^\circ = 0 \quad \text{and} \quad h(x^\circ) = 0$$

and for all $y \in \mathcal{T}(x^\circ) = \{y \mid Dh(x^\circ)y = 0\}$, we have

$$y^T \nabla_x^2 L(x^\circ, \lambda^\circ) y \geq 0$$

Sufficient conditions: if there exists points x° and λ° such that

$$\nabla_x L(x^\circ, \lambda^\circ) = \nabla f(x^\circ) + \sum_{i=1}^{m} \nabla h_i(x^\circ)\lambda_i^\circ = 0 \quad \text{and} \quad h(x^\circ) = 0$$

and for all $y \in \mathcal{T}(x^\circ) = \{y \mid Dh(x^\circ)y = 0\}$, $y \neq 0$, we have

$$y^T \nabla_x^2 L(x^\circ, \lambda^\circ) y > 0$$

then, x° is a (strict) local minimizer

Example

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

- the Lagrangian is

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

- the first-order necessary conditions are

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

$$x_1 + x_2 + x_3 = 3$$

and the solution is $\hat{x} = (1, 1, 1)$, $\lambda = -2$, so $\hat{x}$ is a candidate minimizer we now look at the second-order condition

- note that $\nabla h(x) = (1, 1, 1)$ and the Hessian

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

- however, on the tangent space

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

we have

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

thus, $\hat{x} = (1, 1, 1)$ is not a local minimizer (it is a local maximizer)

Quadratic objective and constraint

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

- $Q = Q^T$ and $P = P^T \succ 0$
- the Lagrangian is

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

- the Lagrange optimality conditions are

$$\begin{aligned} \nabla_x L(x, \lambda) &= 2Qx - 2\lambda P x = 0 \\ \nabla_\lambda L(x, \lambda) &= 1 - x^T P x = 0 \end{aligned}$$

- from the first equation, we have

$$P^{-1}Qx = \lambda x$$

so, optimal points $\hat{x}$ and $\hat{\lambda}$ if they exists, are eigenvectors/eigenvalues of $P^{-1}Q$

- multiplying $P^{-1}Qx = \lambda x$ on the left by x^TP and using $x^TPx = 1$, we get

$$\lambda = x^TQx = f(x)$$

- hence, $f(x) = x^TQx$ is minimized when λ is the smallest eigenvalue of $P^{-1}Q$

- solution x is the eigenvector associated with smallest eigenvalue

Example

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

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

- 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

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

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

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

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

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

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

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

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

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

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

Outline

- equality constrained problems
- **constrained quadratic problems**
- inequality constrained problems
- projected gradient descent

Quadratic program with linear constraints

$$
\begin{array}{ll}
\text{minimize} & (1/2)x^T Q x + r^T x \\
\text{subject to} & Cx = d
\end{array}
$$

- Q is an $n \times n$ symmetric matrix; r is an n-vector
- C is a $p \times n$ matrix; d is a p-vector
- the Lagrangian for this problem is

$$L(x, \lambda) = (1/2)x^T Q x + r^T x + \lambda^T (Cx - d)$$

Optimality conditions

if $x^\star$ is a solution, then there exists $\lambda^\star$ such that:

$$\nabla_x L(x^\star, \lambda^\star) = Qx^\star + r + C^T\lambda^\star = 0 \tag{8.4a}$$

$$Cx^\star - d = 0 \tag{8.4b}$$

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

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

- 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
- conditions are called *KKT optimality conditions*; matrix on left is called *KKT matrix*

Closed-form solution

Assumptions: Q is invertible and C has linearly independent rows

Closed-form solution

- multiply the first equation in (8.4) by Q^{-1} on the left

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

- substituting into the second equation, we get

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

  hence

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

- putting it all together, we get

$$x^\star = Q^{-1}C^T(CQ^{-1}C^T)^{-1}(CQ^{-1}r + d) - Q^{-1}r \tag{8.5}$$

Least norm problem

$$\begin{array}{ll} \text{minimize} & \|x\|^2 \\ \text{subject to} & Cx = d \end{array}$$

- C is a $p \times n$ matrix, d is a p-vector
- the goal is to find the solution of $Cx = d$ with the smallest norm
- a special case of constrained QP with $Q = 2I$ and $r = 0$

Least distance problem: minimizing the distance to a given point $a \neq 0$:

$$\begin{array}{ll} \text{minimize} & \|x - a\|^2 \\ \text{subject to} & Cx = d \end{array}$$

- reduces to least norm problem by a change of variables $y = x - a$

$$\begin{array}{ll} \text{minimize} & \|y\|^2 \\ \text{subject to} & Cy = d - Ca \end{array}$$

- from least norm solution y, we obtain solution $x = y + a$ of first problem

Solution of least norm problem

$$\begin{array}{ll} \text{minimize} & \|x\|^2 \\ \text{subject to} & Cx = d \end{array}$$

Assumption: we assume that C has linearly independent rows

- $Cx = d$ has at least one solution for every d
- C is wide or square $(p \leq n)$; if $p < n$ there are infinite solutions to $Cx = d$

Solution of least norm problem

$$\hat{x} = C^T(CC^T)^{-1}d$$

- solution follows form (8.5) with $Q = 2I$ and $r = 0$
- unique solution under the above assumption
- $C^T(CC^T)^{-1} = C^\dagger$ is the pseudo-inverse of C, which is also a right-inverse

Constrained least squares

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & Cx = d \end{array}$$

- A is an $m \times n$ matrix; b is an m-vector
- C is a $p \times n$ matrix; d is a p-vector
- the objective $\|Ax - b\|^2 = x^T(A^TA)x - 2(A^Tb)^Tx + \|b\|^2$ is quadratic with

$$Q = 2A^TA, \quad r = -2A^Tb$$

- the optimality condition is

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

- since $Q = 2A^TA \succeq 0$, any solution of the above is a global minimizer

Linear quadratic control

Linear dynamical system

$$s_{t+1} = A_t s_t + B_t u_t, \quad y_t = C_t s_t, \quad t = 0, 1, \ldots$$

- n-vector s_t is system *state* (at time t)
- m-vector u_t is system *input* (we control)
- p-vector y_t is system *output*
- s_t, u_t, y_t are typically desired to be small

Objective: choose inputs $u_0, \ldots, u_{T-1}$ that minimizes $J_{\text{output}} + \delta J_{\text{input}}$ with

$$J_{\text{output}} = \|y_0 - y_0^{\text{des}}\|^2 + \cdots + \|y_T - y_T^{\text{des}}\|^2, \quad J_{\text{input}} = \|u_0\|^2 + \cdots + \|u_{T-1}\|^2$$

where y_t^{des} are given desired values (possibly zero)

Constraints

- dynamics constraint
- initial state and (possibly) the final state are specified $s_0 = s^{\text{init}}$, $s_T = s^{\text{des}}$

Linear quadratic control problem

$$
\begin{array}{ll}
\text{minimize} & \|C_0 s_0 - y_0^{\text{des}}\|^2 + \cdots + \|C_T x_T - y_T^{\text{des}}\|^2 + \delta\left(\|u_0\|^2 + \cdots + \|u_{T-1}\|^2\right) \\
\text{subject to} & s_{t+1} = A_t s_t + B_t u_t, \quad t = 0, \ldots, T-1 \\
& s_0 = s^{\text{init}}, \quad s_T = s^{\text{des}}
\end{array}
$$

variables: $s_0, \ldots, s_T$ and $u_0, \ldots, u_{T-1}$

Constrained least squares formulation

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

variables: the $(n(T+1) + mT)$-vector

$$
z = (s_0, \ldots, s_T, u_0, \ldots, u_{T-1})
$$

Linear quadratic control problem

Objective function: $\|\tilde{A}z - \tilde{b}\|^2$ with

$$\tilde{A} = \left[\begin{array}{ccc|ccc} C_0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & C_T & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & \sqrt{\delta} I & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \sqrt{\delta} I \end{array}\right], \quad \tilde{b} = \left[\begin{array}{c} y_0^{\mathrm{des}} \\ \vdots \\ y_T^{\mathrm{des}} \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]$$

Constraints: $\tilde{C}z = \tilde{d}$ with

$$\tilde{C} = \left[\begin{array}{cccccc|cccc} A_0 & -I & 0 & \cdots & 0 & 0 & B_0 & 0 & \cdots & 0 \\ 0 & A_1 & -I & \cdots & 0 & 0 & 0 & B_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{T-1} & -I & 0 & 0 & \cdots & B_{T-1} \\ \hline I & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & I & 0 & 0 & \cdots & 0 \end{array}\right], \quad \tilde{d} = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline s^{\text{init}} \\ s^{\text{des}} \end{array}\right]$$

Linear quadratic regulator

a variation is to consider the linear quadratic control (LQR) objective

$$(1/2)\sum_{t=0}^{T} s_t^T Q_t s_t + (1/2)\sum_{t=0}^{T-1} u_t^T R_t u_t$$

- Q_t and R_t are given matrices of appropriate dimensions
- this problem takes the form:

$$\begin{array}{ll} \text{minimize} & (1/2) z^T \tilde{Q} z \\ \text{subject to} & \tilde{C} z = \tilde{d} \end{array}$$

with the variable $z = (s_0, \ldots, s_T, u_0, \ldots, u_{T-1})$ and the block-diagonal matrix:

$$\tilde{Q} = \left[\begin{array}{ccc|ccc} Q_0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & Q_T & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & R_0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & R_{T-1} \end{array}\right]$$

where $\tilde{C}$ and $\tilde{d}$ are defined as previously specified

Small final state variation

suppose $A_t = A$ and $B_t = B$ and consider the objective:

$$(1/2)\|s_T\|^2 + (1/2)\sum_{t=0}^{T-1} u_t^T R_t u_t$$

- s_T is not predefined but is desired to be small
- it is convenient to iterate the dynamics to express s_T as:

$$s_T = A^T s_0 + Cu$$

  where $u = (u_0, \ldots, u_{T-1})$ and the matrix C is

$$C = \begin{bmatrix} A^{T-1}B & A^{T-2}B & \cdots & AB & B \end{bmatrix}$$

- the control problem then becomes the least norm problem:

$$\begin{array}{ll} \text{minimize} & (1/2)z^T Q z \\ \text{subject to} & [C \quad -I]z = -A^T s_0 \end{array}$$

  with variable $z = (u, s_T)$ and $Q = \mathrm{diag}(R_0, \ldots, R_{T-1}, I)$

Example

consider the discrete-time linear system

$$s_{t+1} = 2s_t + u_t, \quad t \geq 0$$

with $s_0 = 1$; we want to find the values of the inputs u_0 and u_1 that minimizes

$$\frac{1}{2}u_0^2 + \frac{1}{3}u_1^2 + s_2^2$$

- we can formulate this problem as a quadratic program with variables u_0, u_1 and s_2
- the state at time 2 can be found recursively as:

$$s_2 = 2s_1 + u_1 = 2(2s_0 + u_0) + u_2 = 2(2 + u_0) + u_1$$

hence,

$$2u_0 + u_1 - s_2 = -4$$

- therefore, the problem can be formulated as:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}u_0^2 + \frac{1}{3}u_1^2 + s_2^2 \\ \text{subject to} & 2u_0 + u_1 - s_2 = -4 \end{array}$$

- letting $z = (u_0, u_1, s_2)$, we can write the problem as:

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

  where

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

- since Q is invertible and C is a nonzero row vector, the solution is

$$z = (u_0, u_1, s_2) = Q^{-1}C^T(CQ^{-1}C^T)^{-1}d = \left(-\frac{4}{3}, -1, \frac{1}{3}\right)$$

Outline

- equality constrained problems
- constrained quadratic problems
- **inequality constrained problems**
- projected gradient descent

Inequality constrained problems

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{array} \tag{8.6}$$

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

Lagrangian

the *Lagrangian* associated with problem (8.6) is

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

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

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

Regular point

Active inequalities

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

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

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

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

are linearly independent

Motivation of optimality conditions

if x° is a local minimizer of (8.6), then it is a local minimizer of the problem:

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

- applying Lagrange conditions (8.2) to the above problem, we have

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

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

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

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

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

Karush-Kuhn-Tucker (KKT) conditions

if x° is regular and a local minimizer, then there exists $\mu^\circ \in \mathbb{R}^m$, $\lambda^\circ \in \mathbb{R}^p$ such that:

$$\nabla_x L(x^\circ, \mu^\circ, \lambda^\circ) = 0 \tag{8.7a}$$

$$g_i(x^\circ) \le 0, \quad i = 1, \ldots, m \tag{8.7b}$$

$$h_j(x^\circ) = 0, \quad j = 1, \ldots, p \tag{8.7c}$$

$$\mu_i^\circ \ge 0, \quad i = 1, \ldots, m \tag{8.7d}$$

$$\mu_i^\circ g_i(x^\circ) = 0, \quad i = 1, \ldots, m \tag{8.7e}$$

λ° and μ° are called the *Lagrange multiplier* and *KKT multiplier* vectors (or just Lagrange multiplier vectors)

Complementary slackness: condition $\mu_i^\circ g_i(x^\circ) = 0$ implies that

- $g_i(x^\circ) < 0 \Rightarrow \mu_i^\circ = 0$
- $\mu_i^\circ > 0 \Rightarrow g_i(x^\circ) = 0$

called the *complementary slackness*

Example

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

- the Lagrangian is

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

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

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

- to find a solution, assume $\mu_1 = 0$, $x_2 = 0$; then, solving the above, we have

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

  which satisfy the KKT conditions

- for $\mu_2 = 0$, $x_1 = 0$, we get $x_2 = 0$, $\mu_1 = -3$, which violates the condition $\mu \geq 0$

- other combinations $x_1 = x_2 = 0$ and $\mu_1 = \mu_2 = 0$ also violates KKT condition

Necessary conditions: inequality constrained case

Tangent space

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

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

Necessary conditions

suppose x° is regular and a local minimizer, then there exists μ°, λ° such that:

- the KKT conditions (8.7) hold; and
- for all $y \in \mathcal{T}(x^\circ)$, we have

$$y^T \nabla_x^2 L(x^\circ, \mu^\circ, \lambda^\circ) y \geq 0$$

Sufficient conditions: inequality constrained case

Critical tangent space

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

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

Sufficient conditions: suppose there exists points x°, μ°, λ° such that the KKT conditions (8.7) hold and for all $y \in \bar{\mathcal{T}}(x^\circ)$, $y \neq 0$, we have

$$y^T \nabla_x^2 L(x^\circ, \lambda^\circ, \mu^\circ) y > 0$$

then, x° is a strict local minimizer of (8.6)

Example

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

- the Lagrangian is

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

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

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

- 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

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

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

- since both constraints are active, the tangent space is

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

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

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

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

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

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

  this means that the sufficient condition does not hold

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

Example

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

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

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

  are linearly independent

- all feasible points are regular and a local solution must satisfy the KKT conditions

- the Lagrangian is

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

- KKT conditions:

$$\begin{bmatrix} 2x_1 - 2 + \mu - \lambda \\ 1 + \mu + \lambda \end{bmatrix} = 0$$

$$\mu(x_1 + x_2 - 2) = 0$$

$$\mu \geq 0$$

$$x_2 - x_1 - 1 = 0$$

$$x_1 + x_2 - 2 \leq 0$$

- 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 $\hat{x} = (\frac{1}{2}, \frac{3}{2})$ is a local minimizer candidate

- the Hessian of the Lagrangian is

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

for all x (positive semi-definite)

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

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

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

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

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

Outline

- equality constrained problems
- constrained quadratic problems
- inequality constrained problems
- **projected gradient descent**

Projection

the *projection of* $x \in \mathbb{R}^n$ *onto a set* $\mathcal{X} \subseteq \mathbb{R}^n$ is defined as

$$\Pi_{\mathcal{X}}(x) = \underset{z \in \mathcal{X}}{\operatorname{argmin}} \|z - x\|$$

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

Examples

- *box constraint*

$$\mathcal{X} = \{x \mid l_i \leq x_i \leq u_i,\ i = 1, \ldots, n\}, \qquad (\Pi_{\mathcal{X}}(x))_i = \begin{cases} u_i & \text{if } x_i > u_i \\ x_i & \text{if } l_i \leq x_i \leq u_i \\ l_i & \text{if } x_i < l_i \end{cases}$$

- *unit ball constraint:* $\mathcal{X} = \{x \mid \|x\|^2 = 1\}$, $\Pi_{\mathcal{X}}(x) = x/\|x\|$

Gradient descent and projection

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

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

the gradient descent update has the form:

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

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

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

Projected gradient descent

---

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

repeat for $k = 0, 1, \ldots$

1. choose a stepsize α_k
2. update $x^{(k+1)}$:
$$x^{(k+1)} = \Pi_{\mathcal{X}}\big(x^{(k)} - \alpha_k \nabla f(x^{(k)})\big)$$

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

---

$$\Pi_{\mathcal{X}}(x) = \operatorname*{argmin}_{z \in \mathcal{X}} \|z - x\|$$

Examples

- for the problem

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

  the projected gradient descent update is

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

- for the problem

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

  the projected gradient descent update is

$$x^{(k+1)} = (x^{(k)} - \alpha(Qx^{(k)} + r))_+,$$

  where $(\cdot)_+$ replaces negative entries with zero

References and further readings

- E. K.P. Chong, Wu-S. Lu, and S. H. Zak. *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023. (Ch. 20, 21)

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.* Cambridge University Press, 2018. (Ch. 16, 17)