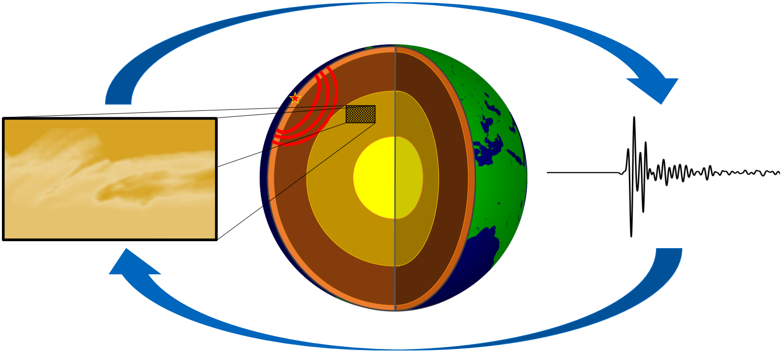
ETH zürich

Short Course on Constrained Nonlinear Optimization

Part I

Christian Boehm

Full-Waveform Inversion



Inverse problem:

$$F(m) = d$$

- m model parameters
- d data
- F forward model (e.g. elastic wave equation)

Inverse problem:

$$F(m) = d$$

- m model parameters
- d data
- F forward model (e.g. elastic wave equation)

Reformulation as optimization problem:

$$\min_m \|F(m) - d\|^2 \quad + \quad \text{regularization}$$

Why additional constraints can be useful:

Why additional constraints can be useful:

- Non-convexity of the problem leads to many local minimizers.

Why additional constraints can be useful:

- Non-convexity of the problem leads to many local minimizers.
- Ill-posedness of the problem requires as much prior information as possible.

Why additional constraints can be useful:

- Non-convexity of the problem leads to many local minimizers.
- Ill-posedness of the problem requires as much prior information as possible.
- Automation of inversion processes.

Constraints on the material can model:

- lower and upper bounds on P- and/or S-wave velocity,
- bounds on the Poisson's ratio,
- restrictions on the total mass,
- positive definiteness of the elastic tensor,
- ...

1 Optimality Conditions

2 Optimization Methods and Algorithms

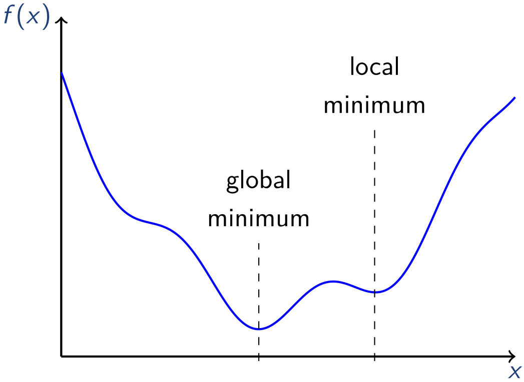
3 Software for Nonlinear Optimization

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X,$$

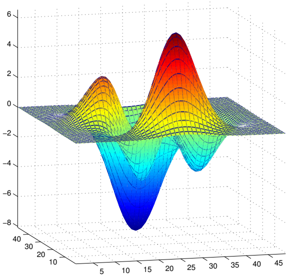
- x optimization variable
- cost function (objective) $f : \mathbb{R}^n \to \mathbb{R}$ and
- admissible set $X \subseteq \mathbb{R}^n$.

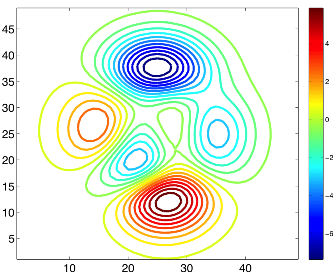
- If $x \in X$ it is called feasible, otherwise infeasible.

Local vs. Global Minima



Peaks


6
4
2
0
−2
−4
−6
−8
40
30
20
10
5
10
15
20
25
30
35
40
45


45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6

We want to find conditions

We want to find conditions

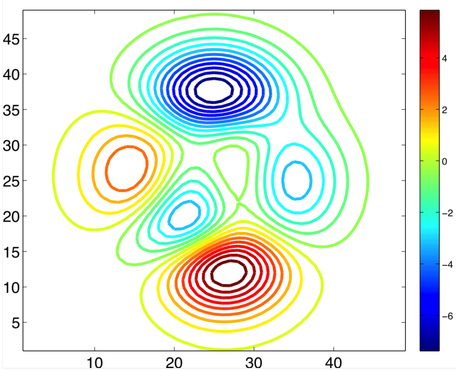
- which are necessary (sufficient?) for local minimizers,

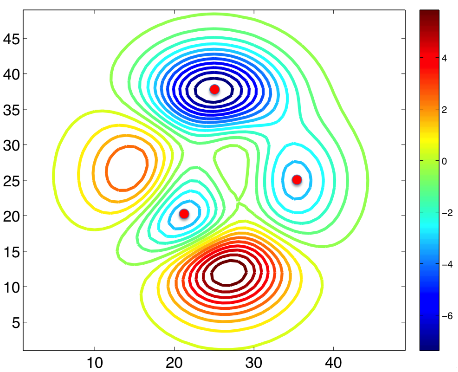
We want to find conditions

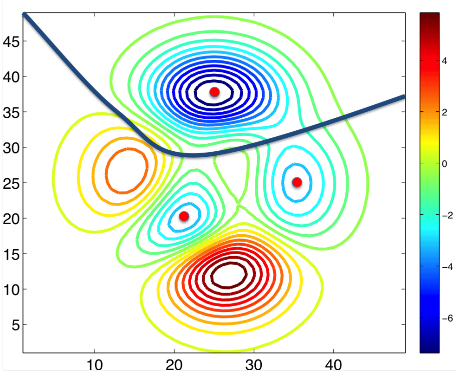
- which are necessary (sufficient?) for local minimizers,
- which can be verified numerically,

We want to find conditions

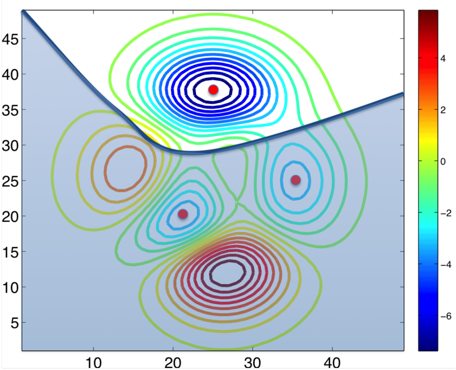
- which are necessary (sufficient?) for local minimizers,
- which can be verified numerically,
- and which are suited for an iterative algorithm.


45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6

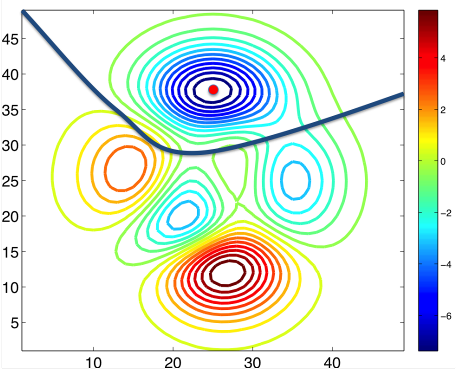

45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6

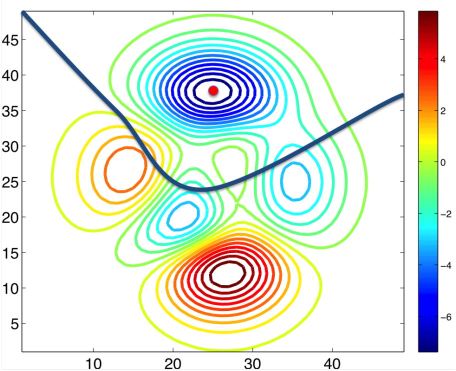

45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6

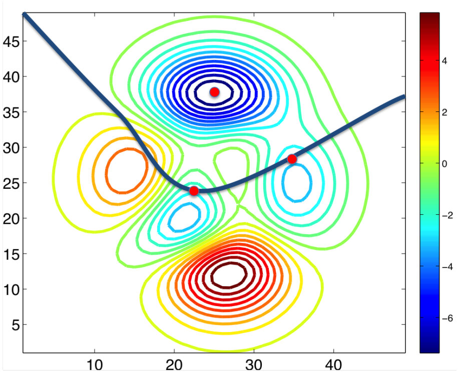
ETH zürich



Example




45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6


45
40
35
30
25
20
15
10
5
10
20
30
40
4
2
0
−2
−4
−6

Characterizing local solutions $\bar{x}$:

$$f(x) \geq f(\bar{x}) \text{ for all } x \text{ in neighborhood of } \bar{x}.$$

Characterizing local solutions $\bar{x}$:

$$f(x) \geq f(\bar{x}) \text{ for all } x \text{ in neighborhood of } \bar{x}.$$

First-order expansion:

$$f(\bar{x} + s) = f(\bar{x}) + \nabla f(\bar{x})^T s + O(\|s\|^2).$$

Characterizing local solutions $\bar{x}$:

$$f(x) \geq f(\bar{x}) \text{ for all } x \text{ in neighborhood of } \bar{x}.$$

First-order expansion:

$$f(\bar{x} + s) = f(\bar{x}) + \nabla f(\bar{x})^T s + O(\|s\|^2).$$

Hence,

$$\nabla f(\bar{x}) = 0.$$

Characterizing local solutions $\bar{x}$:

$$f(x) \geq f(\bar{x}) \text{ for all } x \text{ in neighborhood of } \bar{x}.$$

First-order expansion:

$$f(\bar{x} + s) = f(\bar{x}) + \nabla f(\bar{x})^T s + O(\|s\|^2).$$

Hence,

$$\nabla f(\bar{x}) = 0.$$

gradient = direction of steepest ascent, i.e.,

negative gradient = direction of steepest descent.

Characterizing local solutions $\bar{x}$:

$$f(x) \geq f(\bar{x}) \text{ for all } x \text{ in neighborhood of } \bar{x}.$$

First-order expansion:

$$f(\bar{x} + s) = f(\bar{x}) + \nabla f(\bar{x})^T s + O(\|s\|^2).$$

Hence,

$$\nabla f(\bar{x}) = 0.$$

gradient	=	direction of steepest ascent, i.e.,
negative gradient	=	direction of steepest descent.

If $\nabla f(x) \neq 0$ the objective will decrease in direction $-\nabla f(x)$.

$$\min_{x \in \mathbb{R}^n} f(x)$$

- 1st order necessary condition:

$$\bar{x} \text{ is local solution of } f(x) \quad \Rightarrow \quad \nabla f(\bar{x}) = 0$$

At $\bar{x}$ there is no descent direction (up to first order).

$\bar{x}$ is called stationary point.

$$\min_{x \in \mathbb{R}^n} \; f(x)$$

- 1st order necessary condition:

$$\bar{x} \text{ is local solution of } f(x) \quad \Rightarrow \quad \nabla f(\bar{x}) = 0$$

  At $\bar{x}$ there is no descent direction (up to first order).

  $\bar{x}$ is called stationary point.

- 2nd order necessary condition: $\quad \nabla^2 f(\bar{x}) \succeq 0$

$$\min_{x \in \mathbb{R}^n} f(x)$$

- 1st order necessary condition:

$$\bar{x} \text{ is local solution of } f(x) \quad \Rightarrow \quad \nabla f(\bar{x}) = 0$$

  At $\bar{x}$ there is no descent direction (up to first order).
  $\bar{x}$ is called stationary point.

- 2nd order necessary condition: $\nabla^2 f(\bar{x}) \succeq 0$

- 2nd order sufficient condition:

$$\left.\begin{aligned} \nabla f(\bar{x}) &= 0 \\ \nabla^2 f(\bar{x}) &\succ 0 \end{aligned}\right\} \quad \Rightarrow \quad \bar{x} \text{ is a local solution.}$$

- If $f(x)$ is convex:
  - Every stationary point is a global solution,
  - i.e., there are no maxima or saddle points and
  - 1st order optimality conditions are sufficient for $\bar{x}$ being a global solution

ETH zürich

- If $f(x)$ is convex:
  - Every stationary point is a global solution,
  - i.e., there are no maxima or saddle points   and
  - 1st order optimality conditions are sufficient for $\bar{x}$ being a global solution
- If $f(x)$ is non-convex:
  - Stationary points may also be maxima or saddle points
  - Many local solutions $\bar{x}$ with different $f(\bar{x})$ might exist

ETH zürich

- If $f(x)$ is convex:
  - Every stationary point is a global solution,
  - i.e., there are no maxima or saddle points and
  - 1st order optimality conditions are sufficient for $\bar{x}$ being a global solution
- If $f(x)$ is non-convex:
  - Stationary points may also be maxima or saddle points
  - Many local solutions $\bar{x}$ with different $f(\bar{x})$ might exist

Usually, numerical methods for finding a local minimum

- seek point $\bar{x}$ with $\nabla f(\bar{x}) = 0$,
- try to avoid convergence to local maxima,
- neglect second order conditions.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

- This condition is called variational inequality.

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T(s-\bar{x}) \geq 0 \quad \text{for all } s \in X.$$

- This condition is called variational inequality.
- Geometric interpretation:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

- This condition is called variational inequality.
- Geometric interpretation:
  - (i) f must not decrease in all feasible directions

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

- This condition is called variational inequality.
- Geometric interpretation:
  - (i) f must not decrease in all feasible directions
  - (ii) the angle between the gradient $f(\bar{x})$ and every feasible directions is less than or equal to 90°.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

Additional assumption: X is a convex set.

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

- This condition is called variational inequality.
- Geometric interpretation:
  - (i) f must not decrease in all feasible directions
  - (ii) the angle between the gradient $f(\bar{x})$ and every feasible directions is less than or equal to 90°.
- If $\bar{x}$ is in the strict interior of X the condition is equivalent to

$$\nabla f(\bar{x})^T s \geq 0 \quad \text{for all } s \in \mathbb{R}^n \quad \Leftrightarrow \quad \nabla f(\bar{x}) = 0.$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers,
- which can be verified numerically,
- and which are suited for an iterative algorithm.

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T(s-\bar{x}) \geq 0 \quad \text{for all } s \in X.$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically,
- and which are suited for an iterative algorithm.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✗
- and which are suited for an iterative algorithm.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✗
- and which are suited for an iterative algorithm. ?

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

- Typically, X can be described as the intersection of hyper-curves and/or half-spaces, i.e.,

$$X := \{x \in \mathbb{R}^n : \quad g_i(x) \leq 0, \quad i = 1, \ldots, m, \\ h_i(x) = 0, \quad i = 1, \ldots, p\}.$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

- Typically, X can be described as the intersection of hyper-curves and/or half-spaces, i.e.,

$$X := \{x \in \mathbb{R}^n : \quad g_i(x) \leq 0, \quad i = 1, \dots, m, \\ h_i(x) = 0, \quad i = 1, \dots, p\}.$$

- With $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ we can rewrite the problem as

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0.$$

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

- Typically, X can be described as the intersection of hyper-curves and/or half-spaces, i.e.,

$$X := \{x \in \mathbb{R}^n : \quad g_i(x) \leq 0, \quad i = 1, \ldots, m, \\ h_i(x) = 0, \quad i = 1, \ldots, p\}.$$

- With $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ we can rewrite the problem as

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0.$$

- Assumption throughout this talk: f, g, h are smooth.

- Constraint on the total mass:

$$\int_{\Omega} \rho(\xi)\, d\xi - m_{\text{total}} = 0.$$

- Constraint on the total mass:

$$\int_\Omega \rho(\xi)\, d\xi - m_{\text{total}} = 0.$$

- Lower and upper bounds on v_p, v_s:

$$v_p^{\min}(\xi) \le v_p(\xi) \le v_p^{\max}(\xi), \quad v_s^{\min}(\xi) \le v_s(\xi) \le v_s^{\max}(\xi).$$

The problem is called convex,

- if f is a convex function and X is a convex set or
- if f and g are convex functions and h is affine linear.

ETH zürich

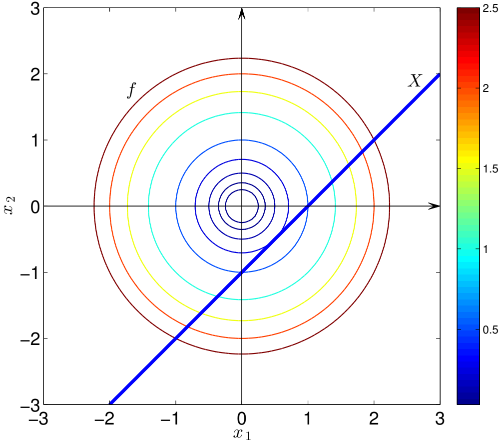
The problem is called convex,

- if f is a convex function and X is a convex set or
- if f and g are convex functions and h is affine linear.

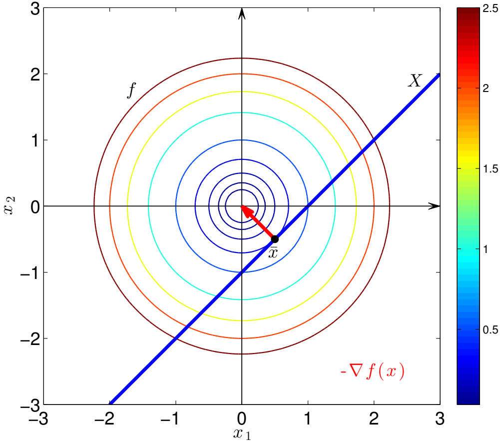
Similar properties as in the unconstrained case:

- Every stationary point is a global solution.
- There are no maxima or saddle points.
- 1st order optimality conditions are sufficient for $\bar{x}$ being a global solution.

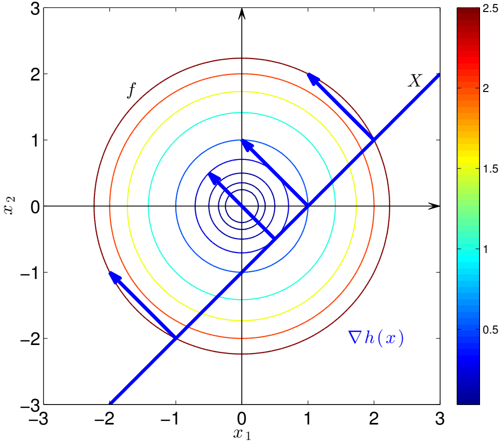
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



$$\min_{x \in \mathbb{R}^n} \ f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$

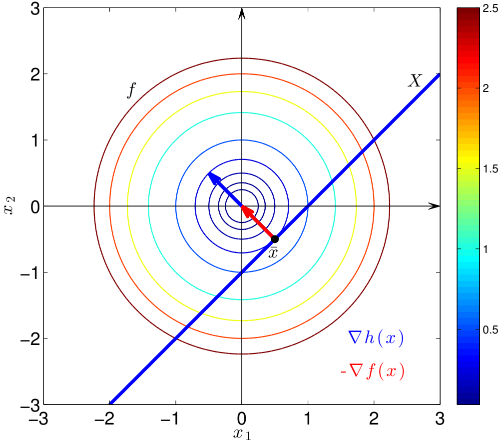


$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$

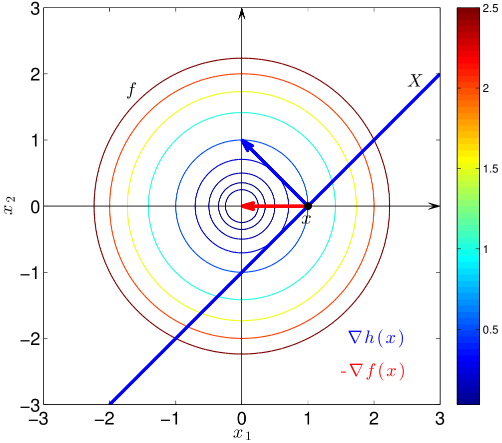


ETH zürich

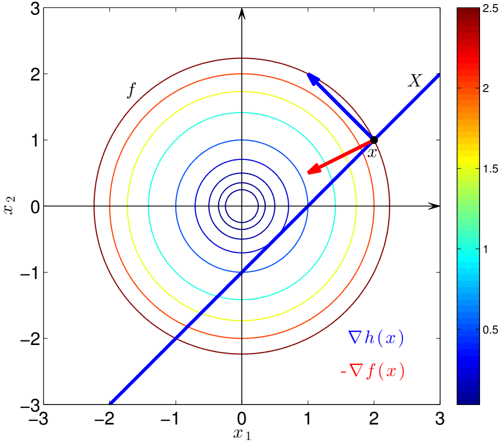
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



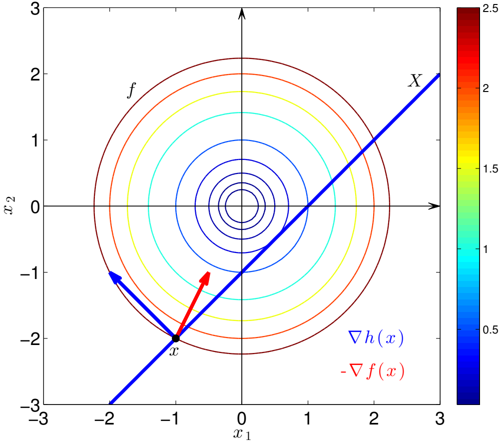
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



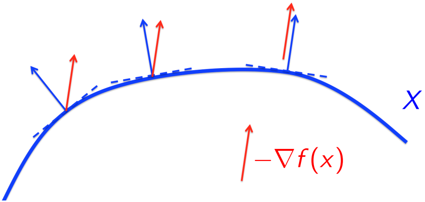
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



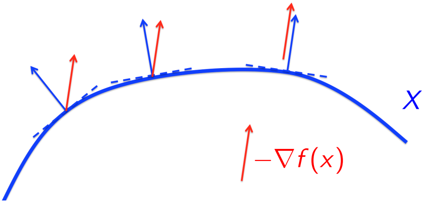
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$

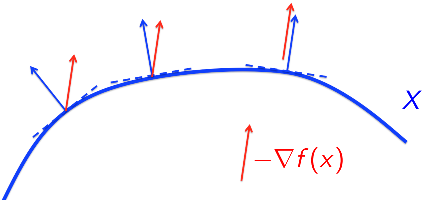


$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \tag{P}$$



- For feasible x, $\nabla h(x)$ is orthogonal to $\{x \in \mathbb{R}^n : h(x) = 0\}$.

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$



- For feasible x, $\nabla h(x)$ is orthogonal to $\{x \in \mathbb{R}^n : h(x) = 0\}$.
- First order optimality conditions:
  There exists $\bar{\mu} \in \mathbb{R}^p$ such that

$$\sum_{i=1}^{p} \bar{\mu}_i \nabla h_i(\bar{x}) = -\nabla f(\bar{x})$$

$$h(\bar{x}) = 0.$$

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \tag{P}$$

- Assumptions:
  - $\bar{x}$ is a local solution of (P)
  - constraint qualification: $\nabla h_1(\bar{x}), \ldots, \nabla h_p(\bar{x})$ are linearly independent
- Then there exists $\bar{\mu} \in \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad h(x) = 0. \tag{P}$$

- Assumptions:
  - $\bar{x}$ is a local solution of (P)
  - constraint qualification: $\nabla h_1(\bar{x}), \dots, \nabla h_p(\bar{x})$ are linearly independent
- Then there exists $\bar{\mu} \in \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

- Lagrangian function: $L(x, \mu) := f(x) + h(x)^T \mu$

$$\begin{aligned} \nabla_x L(\bar{x}, \bar{\mu}) &= 0, \\ \nabla_\mu L(\bar{x}, \bar{\mu}) &= 0. \end{aligned}$$

$\rightarrow$ $\bar{\mu}$ is called (optimal) Lagrangian multiplier

We have to assume the constraint qualification, otherwise:

$$\min_{x \in \mathbb{R}^2} f(x) := (x_1 - 1)^2 + (x_2 + 1)^2 \quad \text{s.t.} \quad h(x) := x_1^2 = 0.$$

We have to assume the constraint qualification, otherwise:

$$\min_{x\in\mathbb{R}^2} f(x) := (x_1 - 1)^2 + (x_2 + 1)^2 \quad \text{s.t.} \quad h(x) := x_1^2 = 0.$$

Then:

$$\bar{x} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla f(\bar{x}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad \nabla h(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Obviously, Lagrangian multiplier $\bar{\mu}$ does not exist.

We want to find conditions

- which are necessary (sufficient?) for local minimizers,
- which can be verified numerically,
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

This is a system of nonlinear equations.

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically,
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

This is a system of nonlinear equations.

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✔
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

This is a system of nonlinear equations.

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✔
- and which are suited for an iterative algorithm. ✔

Find $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0. \end{aligned}$$

This is a system of nonlinear equations.

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

For every m there is a unique solution $u(m)$ to $\mathcal{L}(u,m) = f$.

Equivalent unconstrained problem

$$\min_{m} \; j(m) := \chi(u(m))$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

For every m there is a unique solution $u(m)$ to $\mathcal{L}(u,m) = f$.

Equivalent unconstrained problem

$$\min_{m} \; j(m) := \chi(u(m))$$

Optimality conditions:

$$0 = \nabla j(m) = \nabla_m \mathcal{L}(u(m), m) \cdot u^\dagger$$

with adjoint field $u^\dagger$ determined by

$$\nabla_u \mathcal{L}(u(m), m) u^\dagger = -\nabla \chi(u(m)).$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Define $x := (u, m)^T, \quad f(x) := \chi(u), \quad h(x) := \mathcal{L}(u, m) - F$.

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Define $x := (u, m)^T$, $\quad f(x) := \chi(u)$, $\quad h(x) := \mathcal{L}(u,m) - F$.

Optimality conditions:

$$\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} = 0$$

$$h(\bar{x}) = 0$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Define $x := (u,m)^T$, $f(x) := \chi(u)$, $h(x) := \mathcal{L}(u,m) - F$.

Optimality conditions:

$$\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} = 0$$

$$h(\bar{x}) = 0 \quad \Leftrightarrow \quad \mathcal{L}(\bar{u},\bar{m}) = F.$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Define $x := (u,m)^T$, $\quad f(x) := \chi(u)$, $\quad h(x) := \mathcal{L}(u,m) - F$.

Optimality conditions:

$$\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} = 0 \quad \Leftrightarrow \quad \begin{array}{ccccc} \nabla\chi(\bar{u}) & + & \nabla_u\mathcal{L}(\bar{u},\bar{m})\bar{\mu} & = & 0 \\ 0 & + & \nabla_m\mathcal{L}(\bar{u},\bar{m})\bar{\mu} & = & 0 \end{array}$$

$$h(\bar{x}) = 0 \quad \Leftrightarrow \quad \mathcal{L}(\bar{u},\bar{m}) = F.$$

Seismic Tomography:

$$\min_{u,m} \underbrace{\chi(u)}_{\text{misfit}} \quad \text{s.t.} \quad \underbrace{\mathcal{L}(u,m) = F}_{\text{elastic wave equation}}$$

Define $x := (u,m)^T$, $\quad f(x) := \chi(u)$, $\quad h(x) := \mathcal{L}(u,m) - F$.

Optimality conditions:

$$\nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\mu} = 0 \quad \Leftrightarrow \quad \begin{array}{ccccc} \nabla\chi(\bar{u}) & + & \nabla_u \mathcal{L}(\bar{u},\bar{m})\bar{\mu} & = & 0 \\ 0 & + & \nabla_m \mathcal{L}(\bar{u},\bar{m})\bar{\mu} & = & 0 \end{array}$$

$$h(\bar{x}) = 0 \quad \Leftrightarrow \quad \mathcal{L}(\bar{u},\bar{m}) = F.$$

Adjoint state and Lagrangian multiplier are the same!

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \tag{P}$$

Lagrangian function: $L(x, \mu) := f(x) + h(x)^T \mu$, same assumptions as before.

- 2nd order necessary conditions:
  If $\bar{x}$ is local minimum, then for all $d \in \mathbb{R}^n$ with $\nabla h(\bar{x})^T d = 0$:

$$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu})^T d \geq 0$$

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$

Lagrangian function: $L(x, \mu) := f(x) + h(x)^T \mu$, same assumptions as before.

- 2nd order necessary conditions:
  If $\bar{x}$ is local minimum, then for all $d \in \mathbb{R}^n$ with $\nabla h(\bar{x})^T d = 0$:

  $$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu})^T d \geq 0$$

- 2nd order sufficient conditions:
  If $\bar{x}$ satisfies 1st order necessary conditions and for all $d \in \mathbb{R}^n \setminus \{0\}$ with $\nabla h(\bar{x})^T d = 0$:

  $$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu})^T d > 0,$$

  then $\bar{x}$ is local a solution.

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad h(x) = 0. \qquad \text{(P)}$$

Lagrangian function: $L(x, \mu) := f(x) + h(x)^T \mu$, same assumptions as before.

- 2nd order necessary conditions:
  If $\bar{x}$ is local minimum, then for all $d \in \mathbb{R}^n$ with $\nabla h(\bar{x})^T d = 0$:

  $$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu})^T d \geq 0$$

- 2nd order sufficient conditions:
  If $\bar{x}$ satisfies 1st order necessary conditions and for all $d \in \mathbb{R}^n \setminus \{0\}$ with $\nabla h(\bar{x})^T d = 0$:

  $$d^T \nabla^2_{xx} L(\bar{x}, \bar{\mu})^T d > 0,$$

  then $\bar{x}$ is local a solution.

  $\rightarrow$ "Projection" of $\nabla^2_{xx} L(\bar{x}, \bar{\mu})$ onto X is positive (semi-)definite

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$

- g_i is called active in $x \in X$ if $g_i(x) = 0$ and inactive if $g_i(x) < 0$.

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$

- g_i is called active in $x \in X$ if $g_i(x) = 0$ and inactive if $g_i(x) < 0$.
- Only active constraints influence the set of feasible regions.

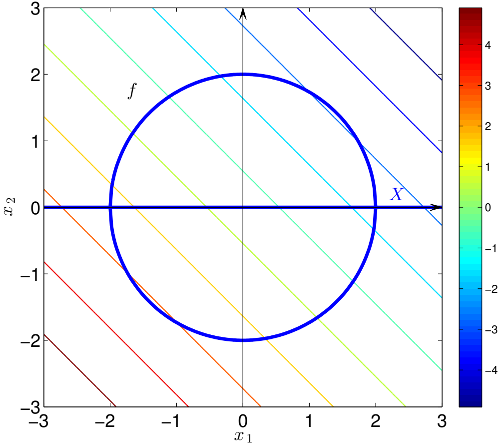
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$

- g_i is called active in $x \in X$ if $g_i(x) = 0$ and inactive if $g_i(x) < 0$.
- Only active constraints influence the set of feasible regions.

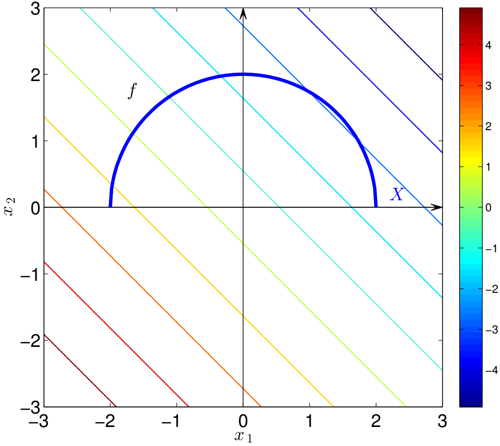
Example:

$$X = \{x \in \mathbb{R}^2 : \|x\| = 2, \quad -x_2 \leq 0\}$$

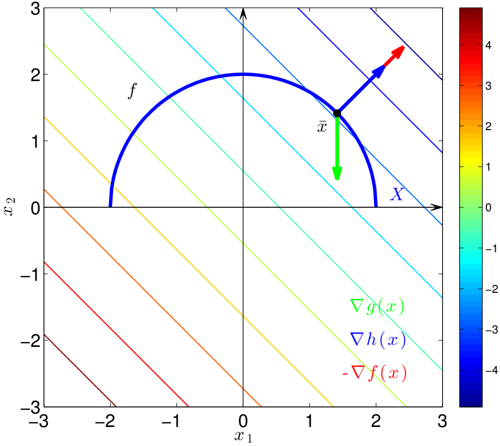
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$



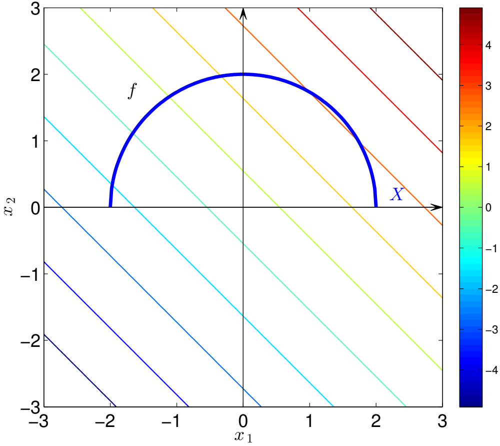
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$



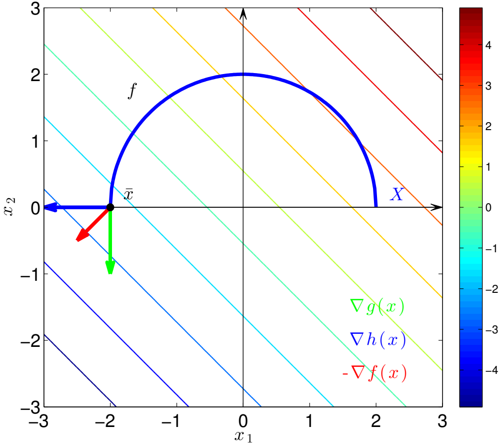
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad g(x) \le 0, \quad h(x) = 0. \qquad \text{(P)}$$



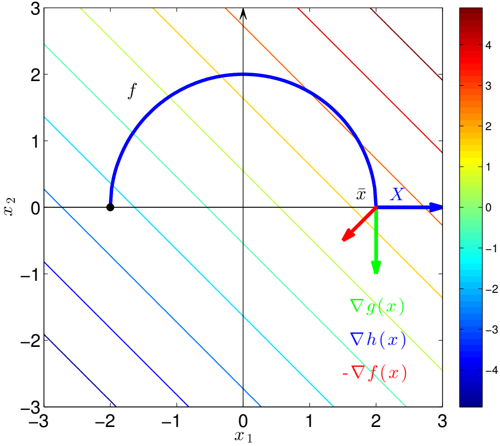
$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$



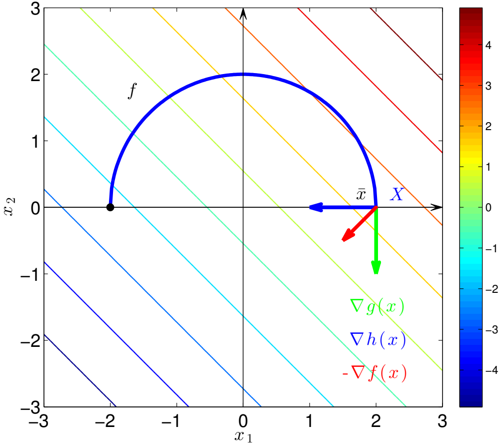
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$



$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$



$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$



$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \qquad \text{(P)}$$

Karush-Kuhn-Tucker conditions:

There exist $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$

Karush-Kuhn-Tucker conditions:

There exist $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

- necessary optimality conditions if a constraint qualification holds, e.g.

$$\nabla h_1(\bar{x}), \dots, \nabla h_p(x), \nabla g_i(x) \ \text{ for } g_i \text{ active, are linearly independent}$$

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0. \tag{P}$$

Karush-Kuhn-Tucker conditions:

There exist $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

- necessary optimality conditions if a constraint qualification holds, e.g.

  $\nabla h_1(\bar{x}), \ldots, \nabla h_p(x), \nabla g_i(x)$ for g_i active, are linearly independent

- sufficient(!) conditions for convex problems

We want to find conditions

- which are necessary (sufficient?) for local minimizers,
- which can be verified numerically,
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically,
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✔
- and which are suited for an iterative algorithm.

Find $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ such that

$$\begin{aligned} \nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} &= 0, \\ h(\bar{x}) &= 0, \\ g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) &= 0. \end{aligned}$$

We want to find conditions

- which are necessary (sufficient?) for local minimizers, ✔
- which can be verified numerically, ✔
- and which are suited for an iterative algorithm. ?

Find $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ such that

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{\lambda} + \nabla h(\bar{x})\bar{\mu} = 0,$$
$$h(\bar{x}) = 0,$$
$$g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) = 0.$$

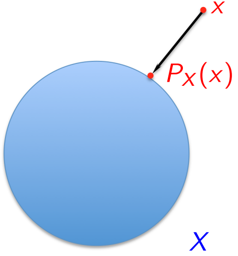
$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in X$$

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})^T (s - \bar{x}) \geq 0 \quad \text{for all } s \in X.$$

For convex X this is equivalent to

$$\bar{x} = P_X(\bar{x} - \gamma \nabla f(\bar{x}))$$

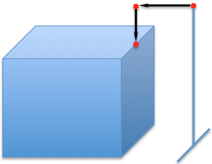
where P_X is the projection onto X and $\gamma > 0$.



$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad x^l \leq x \leq x^u$$

Then

$$P_X(x)_i = \begin{cases} x_i^l & \text{if } x_i \leq x_i^l, \\ x_i & \text{if } x_i^l < x_i < x_i^u, \\ x_i^u & \text{if } x_i \geq x_i^u. \end{cases}$$



Optimality conditions:

$$\bar{x} \in X, \quad \nabla f(\bar{x})_i \begin{cases} \geq 0 & \text{if } \bar{x}_i = x_i^l, \\ = 0 & \text{if } x_i^l < \bar{x}_i \leq x_i^u, \\ \leq 0 & \text{if } \bar{x}_i = x_i^u. \end{cases}$$

Two types of optimality conditions:

- projection formula

  $\rightarrow$ involves projection onto the feasible set

  and a system of nonlinear equations.

- KKT conditions

  $\rightarrow$ necessary conditions if a constraint qualification is satisfied

  $\rightarrow$ involves Lagrangian multiplier

  and a system of nonlinear equations ($+$ sign constraints)

Both are sufficient for convex problems.

Not covered here:

- constraint qualifications
- non-smooth functions, i.e., if derivatives are not available
- optimization problems in function space
- globalization strategies (i.e. line search or trust region methods)

- Books on nonlinear optimization:
  - J. Nocedal and S. J. Wright: Numerical Optimization (2nd edition), Springer 2006
  - C. T. Kelley: Iterative Methods for Optimization. SIAM 1999
  - M. Hinze, R. Pinnau, S. Ulbrich, and M. Ulbrich: Optimization with PDE Constraints. Springer 2009

- Websites with optimization codes
  - Decision Tree of Optimization Software: `http://plato.la.asu.edu/guide.html`
  - NEOS Guide: `http://www.neos-guide.org/Optimization-Guide`