Introduction to Nonlinear Programming[1]

Sébastien Gros

OPTEC, KU Leuven

Workshop on Modern Nonparametric Methods for Time Series, Reliability & Optimization

[1] A part of the material was provided by B. Chachuat, Imperial College, b.chachuat@imperial.ac.uk

Outline

Some definitions

Optimization problem:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \end{array}$$

where: $\mathbf{x} \in \mathbb{R}^n, f \in \mathbb{R}, \mathbf{g} \in \mathbb{R}^{m_i}, \mathbf{h} \in \mathbb{R}^{m_e}$

Basic terminology:

- Function $f \in \mathbb{R}$ is the cost/objective/penalty function
- Functions $\mathbf{h}$ and $\mathbf{g}$ yields the equality & inequality constraints, resp.
- The set of indices i for which $g_i(\mathbf{x}) = 0$ is named the active set

Some classes of problems:

- Convex NLP: $f, \mathbf{g}$ convex, $\mathbf{h}$ affine
- Non-smooth: any of $f, \mathbf{g}, \mathbf{h}$ is not $\mathcal{C}^1$
- (Mixed)-Integer Programming: some $\mathbf{x}$ take only discrete values

Defining (Global) Optimality

Feasible Set

The feasible set S (or feasible region) of an optimization model is the collection of choices for decision variables satisfying all model constraints: $S \triangleq \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}) = \mathbf{0}\}$. Any point $\bar{\mathbf{x}} \in S$ is said feasible.

(Global) Optimum

An optimal solution, $\mathbf{x}^*$, is a feasible point with objective function value lower than any other feasible point, i.e. $\mathbf{x}^* \in S$ and $f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in S$

1. The optimal value f^* in an optimization model is the objective function value of any optimal solutions: $f^* = f(\mathbf{x}^*)$ — It is unique!
2. But, an optimization model may have:
   - a unique optimal solution
   - several alternative optimal solutions
   - no optimal solutions (unbounded or infeasible models)

Defining Local Optimality

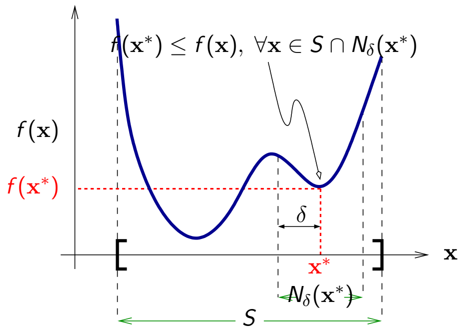
Local Optimum

A point $\mathbf{x}^*$ is a local optimum for the function $f : \mathbb{R}^n \to \mathbb{R}$ on the set S if it is feasible ($\mathbf{x}^* \in S$) and if sufficiently small neighborhoods surrounding it contain no points that are both feasible and lower in objective value:

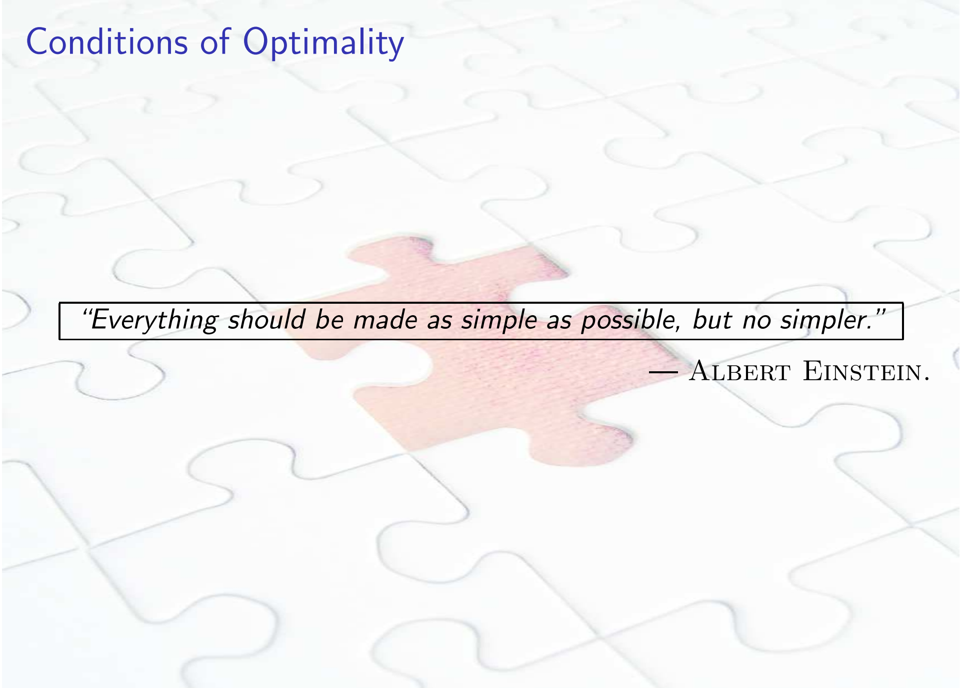
$$\exists \delta > 0 : \quad f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \in S \cap B_\delta(\mathbf{x}^*)$$

Remarks:

1. Global optima are always local optima
2. Local optima may not be global optima



Conditions of Optimality



"Everything should be made as simple as possible, but no simpler."

— Albert Einstein.

Algebraic Characterization of Unconstrained Local Optima

1st-Order **Necessary** Condition of Optimality (FONC)

$$\mathbf{x}^* \text{ local optimum} \quad \Rightarrow \quad \nabla f(\mathbf{x}^*) = \mathbf{0},\ \mathbf{x}^* \text{ stationary point}$$

2nd-Order **Sufficient** Conditions of Optimality (SOSC)

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \text{ and } \nabla^2 f(\mathbf{x}^*) \succeq 0 \quad \Rightarrow \quad x^* \text{ strict local minimum}$$

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \text{ and } \nabla^2 f(\mathbf{x}^*) \preceq 0 \quad \Rightarrow \quad x^* \text{ strict local maximum}$$

No conclusion can be drawn in case $\nabla^2 f(\mathbf{x}^*)$ is indefinite!

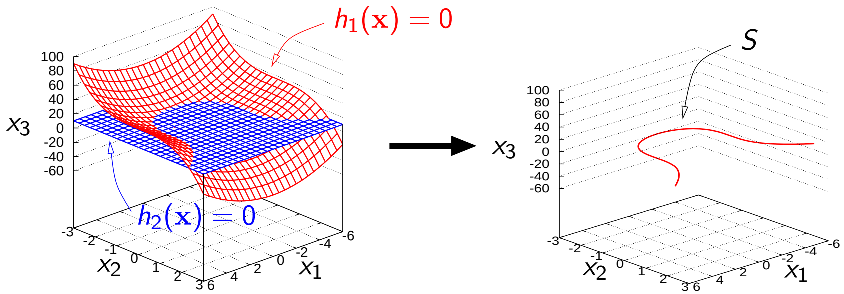
Remarks:

- Stationarity of f: $\nabla f(\mathbf{x}^*) = 0 \Leftrightarrow \nexists\, \mathbf{d}$ such that $\nabla f(\mathbf{x}^*)^T \mathbf{d} < 0$
- Positivity of $\nabla^2 f \Leftrightarrow \forall\, \mathbf{d} \neq 0,\ \mathbf{d} \nabla^2 f(\mathbf{x}^*) \mathbf{d}^T > 0$

General Equality Constraints

- The intersection of the equality constraints, $h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})$ forms the feasible set: $S \triangleq \{\mathbf{x} \in \mathbb{R}^n : h_1(\mathbf{x}) = \cdots = h_m(\mathbf{x}) = 0\} \subsetneq \mathbb{R}^n$
- A feasible point $\bar{\mathbf{x}} \in S \triangleq \{\mathbf{x} : h_i(\mathbf{x}) = 0, i = 1, \ldots, m_e\}$ is called regular if rank $\left(\nabla h_1(\bar{\mathbf{x}}) \quad \ldots \quad \nabla h_{m_e}(\bar{\mathbf{x}}) \right) = m_e$

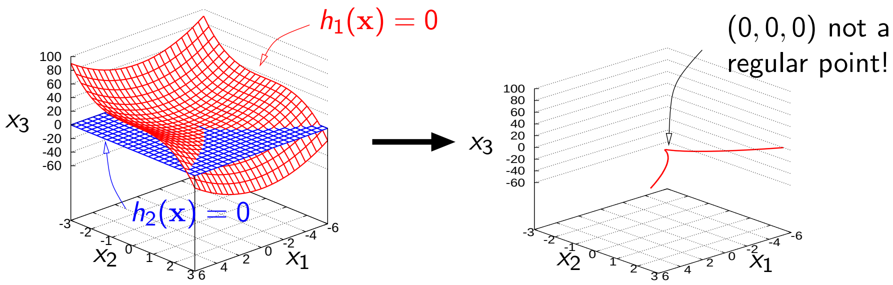
Example. Consider the constraints $h_1(\mathbf{x}) \triangleq x_1^2 - 2x_2^3 - x_3 = 0$ and $h_2(\mathbf{x}) \triangleq x_3 - 10 = 0$



Tangent Set to General Equality Constraints

- Consider the collection of *all* smooth curves in S passing through a point $\bar{\mathbf{x}} \in S$. The tangent set of S at $\bar{\mathbf{x}}$, denoted by $\mathscr{T}S\bar{\mathbf{x}}$, is the collection of the vectors tangent to *all* these curves at $\bar{\mathbf{x}}$
- At a regular point $\bar{\mathbf{x}} \in S$, $\mathscr{T}S\bar{\mathbf{x}} \triangleq \{\mathbf{d} \in \mathbb{R}^n : \nabla h_j(\bar{\mathbf{x}})^\mathsf{T}\mathbf{d} = 0, \forall j\}$

Example. Consider the constraints $h_1(\mathbf{x}) \triangleq x_1^2 - 2x_2^3 - x_3 = 0$ and $h_2(\mathbf{x}) \triangleq x_3 = 0$



Geometric Optimality Condition

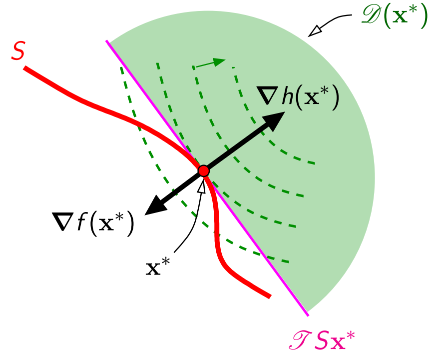
A point $\mathbf{x}^* \in \mathbb{R}^n$ is a local optimum of a real-value function f in the feasible set $S \subset \mathbb{R}^n$, if sufficiently small neighborhoods surrounding it contain no feasible points that are lower in objective value

Geometric Optimality Condition

Let f in $\mathcal{C}^1$, and let $S \triangleq \{\mathbf{x} : h_1(\mathbf{x}) = \cdots = h_m(\mathbf{x}) = 0\}$. If $\mathbf{x}^*$ is an optimum for f on S, then

$$\mathscr{D}(\mathbf{x}^*) \cap \mathscr{T}S\mathbf{x}^* = \emptyset,$$

with:
- $\mathscr{D}(\mathbf{x}^*)$, set of improving directions
- $\mathscr{T}S\mathbf{x}^*$, tangent set



$$\nexists \mathbf{d} \in \mathbb{R}^n : \nabla f(\mathbf{x}^*)^\mathsf{T}\mathbf{d} < 0 \text{ and } \nabla h_i(\mathbf{x}^*)^\mathsf{T}\mathbf{d} = 0$$

... only if $\mathbf{x}^*$ is regular!

FONC for equality constraints

First-order Necessary Conditions

Let $f, h_1, \ldots, h_m$ in $\mathcal{C}^1$. If $\mathbf{x}^*$ is a (local) optimum for f s.t. $h_i(\mathbf{x}) = 0$, $i = 1, \ldots, m$, **and** $\mathbf{x}^*$ regular, there is a unique vector $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:

$$\begin{cases} h_1(\mathbf{x}^*) = \cdots = h_m(\mathbf{x}^*) = 0 \\ \nabla f(\mathbf{x}^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(\mathbf{x}^*) \end{cases}$$

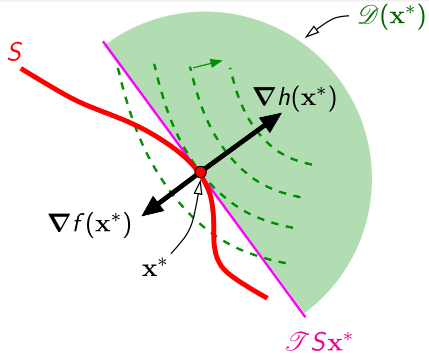
Square system: $(n + m)$ conditions in $(n + m)$ variables $(\mathbf{x}, \boldsymbol{\lambda})$

Lagrange multipliers: $\lambda_i \leftrightarrow h_i$

Lagrangian stationarity:

$$\nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \triangleq f(\mathbf{x}) - \boldsymbol{\lambda}^\mathsf{T}\mathbf{h}(\mathbf{x})$ is called the Lagrangian



Interpretation of the Lagrange Multipliers

Optimization Problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{b}$$

FONC: Find $\mathbf{x}^* \in \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\begin{cases} h_j(\mathbf{x}^*) - b_j = 0, \quad \forall j = 1, \ldots, m \\ \nabla f(\mathbf{x}^*) - \sum_{i=1}^m \lambda_i^* \nabla h_i(\mathbf{x}^*) = \mathbf{0} \end{cases}$$

- A variation in RHS b_j affects the optimal solution $\longrightarrow \mathbf{x}^*(b_j), \boldsymbol{\lambda}^*(b_j)$
- Rate of change in optimal solution value w.r.t. b_j:

$$\left.\frac{\partial f(\mathbf{x}^*(b_j))}{\partial b_j}\right|_{b_j=0} = \lambda_j^*$$

The optimal Lagrange multiplier, λ_j^*, associated with constraint $h_j(\mathbf{x}) = b_j$ can be interpreted as the rate of change in optimal value for infinitesimal change in RHS b_j

Inequality Constrained Optimization

- Consider the optimization problem with inequality constraints

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

- In general, multiple inequality constraints define a (possibly unbounded) n-dimensional set,

$$S \triangleq \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \leq 0, i = 1,\dots,m\} \subseteq \mathbb{R}^n$$

- At a feasible point $\bar{\mathbf{x}}$, the ith constraints is said to be active if $g_i(\bar{\mathbf{x}}) = 0$; it is said to be inactive if $g_i(\bar{\mathbf{x}}) < 0$
- The set of active constraints at a feasible point $\bar{\mathbf{x}}$ is

$$\mathcal{A}\bar{\mathbf{x}} \triangleq \{i : g_i(\bar{\mathbf{x}}) = 0\}$$

Arguably more difficult than equality constraints since active set not known a priori...

Characterizing Feasible Directions

- Consider the feasible domain $S \triangleq \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, i = 1,\dots,m\}$
- The set of feasible directions at a point $\bar{\mathbf{x}} \in S$ is

$$\mathscr{F}(\bar{\mathbf{x}}) \triangleq \{\mathbf{d} \neq \mathbf{0} : \exists \varepsilon > 0 \text{ such that } \bar{\mathbf{x}} + \alpha\mathbf{d} \in S, \forall \alpha \in (0,\epsilon)\}$$

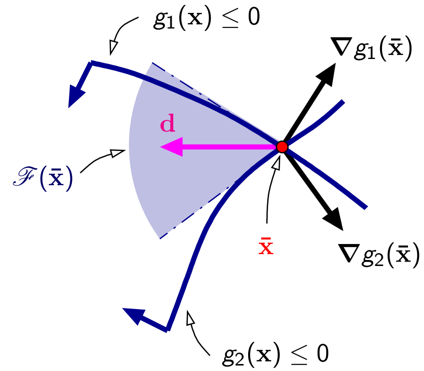
Algebraic Characterization

Let $g_1,\dots,g_m$ in $\mathcal{C}^1$, and let $\bar{\mathbf{x}} \in S$. Any direction $\mathbf{d} \in \mathbb{R}^n$ such that

$$\nabla g_i(\bar{\mathbf{x}})^\mathsf{T}\mathbf{d} < 0, \quad \forall i \in \mathcal{A}\bar{\mathbf{x}}$$

is a feasible direction

- This condition is sufficient, yet not necessary!
- What if $\mathcal{A}\bar{\mathbf{x}} = \emptyset$?



Regularity of General Inequality Constraints

- A point $\bar{\mathbf{x}} \in S$ is said to be regular if the gradient vectors $\nabla g_i(\bar{\mathbf{x}})$ for all the active constraints are linearly independent:

$$\text{rank}\left(\ \nabla g_i(\bar{\mathbf{x}}), i \in \mathcal{A}\bar{\mathbf{x}}\ \right) = \text{card}\left(\mathcal{A}\bar{\mathbf{x}}\right)$$

- Point $\bar{\mathbf{x}} \in S$ is then also said to respect the Linear Independence Constraint Qualification (LICQ).
- When mixing equality constraints with inequality constraints, the active constraints can be seen as equality constraints and the inactive constraints disregarded, i.e. a point respects LICQ iff

$$\text{rank}\left(\ \nabla g_i(\bar{\mathbf{x}}),\ \nabla h_j(\bar{\mathbf{x}}), i \in \mathcal{A}\bar{\mathbf{x}},\ j = 1,...,m_e\ \right) = \text{card}\left(\mathcal{A}\bar{\mathbf{x}}\right)$$

and is said to be regular.

There are more (and less demanding) Constraint Qualifications.

Geometric Optimality Condition

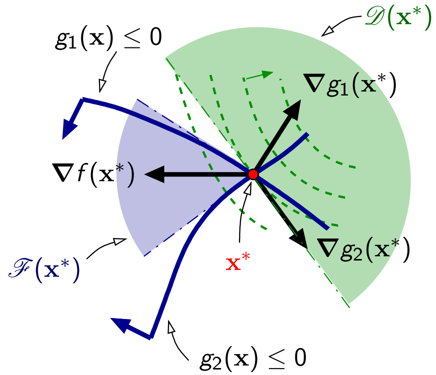
A point $\mathbf{x}^* \in \mathbb{R}^n$ is a local optimum of a real-value function f in the feasible set $S \subset \mathbb{R}^n$, if sufficiently small neighborhoods surrounding it contain no feasible points that are lower in objective value

Geometric Optimality Condition

Let f in $\mathcal{C}^1$, and let $S \triangleq \{\mathbf{x} : g_i(\mathbf{x}) \leq 0, i = 1,\dots,m\}$. If $\mathbf{x}^*$ is an optimum for f on S, then

$$\mathscr{D}(\mathbf{x}^*) \cap \mathscr{F}(\mathbf{x}^*) = \emptyset,$$

with:
- $\mathscr{D}(\mathbf{x}^*)$, set of improving directions
- $\mathscr{F}(\mathbf{x}^*)$, set of feasible directions



$$\nexists \mathbf{d} \in \mathbb{R}^n : \nabla f(\mathbf{x}^*)^\mathsf{T}\mathbf{d} < 0 \ \underline{\text{and}}\ \nabla g_i(\mathbf{x}^*)^\mathsf{T}\mathbf{d} < 0, \forall i \in \mathcal{A}\mathbf{x}^*$$

... only if $\mathbf{x}^*$ is regular!

KKT Points

- Let f and $g_1, \dots, g_m$ in $\mathcal{C}^1$, and consider the NLP problem

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize:}} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1, \dots, m \end{aligned}$$

- A point $(\bar{\mathbf{x}}, \bar{\boldsymbol{\nu}}) \in \mathbb{R}^n \times \mathbb{R}^m$ is called a KKT point if it satisfied

$$\begin{array}{rl} \text{Primal Feasibility:} & g_i(\bar{\mathbf{x}}) \le 0, \quad i = 1, \dots, m \\ \text{Dual Feasibility:} & \begin{cases} \nabla f(\bar{\mathbf{x}}) = \sum_{i=1}^{m} \bar{\nu}_i \nabla g_i(\bar{\mathbf{x}}) \\ \bar{\nu}_i \le 0, \quad i = 1, \dots, m \end{cases} \\ \text{Complementarity Slackness:} & \bar{\nu}_i g_i(\bar{\mathbf{x}}) = 0, \quad i = 1, \dots, m \end{array}$$

FONC for Inequality Constraints

First-Order Necessary Conditions

Let $f, g_1, \dots, g_m$ be $\mathcal{C}^1$. If $\mathbf{x}^*$ is a (local) optimum for f s.t. $g_i(\mathbf{x}) \le 0$, $i = 1, \dots, m$, **and** regular, there is a unique vector $\boldsymbol{\nu}^* \in \mathbb{R}^m$ such that $(\mathbf{x}^*, \boldsymbol{\nu}^*)$ is a KKT point:

$$\begin{array}{ll} \boldsymbol{\nu}^* \le \mathbf{0}, & g(\mathbf{x}^*) \le \mathbf{0}, \\ \nabla \mathcal{L}(\mathbf{x}^*, \nu^*) = \mathbf{0}, & \nu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m \end{array}$$

where $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*) = f(\mathbf{x}^*) - \sum_{i=1}^{m} \nu_i^* g_i(\mathbf{x}^*)$.

KKT Multipliers (Minimize):

- $g_i(\mathbf{x}) \le 0 \leftrightarrow \nu_i^* \le 0$

Interpretation:

- $g_i(\mathbf{x}) \le b_i$ active $\Rightarrow \nu_i^* \triangleq \frac{\partial f^*}{\partial b_i}$

Active Set Selection:

1. Pick-up active set (a priori)
2. Calculate KKT point $(\mathbf{x}^*, \boldsymbol{\nu}^*)$ (if any)

$\rightarrow$ Repeat for **all** possible active sets!

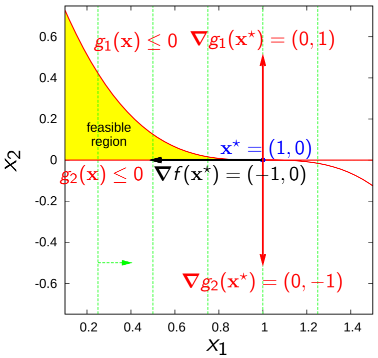
Solving Nonlinear Inequality Constrained Optimization

A simple problem where $\mathbf{x}^*$ is not a KKT point...

$$\min_{\mathbf{x}} \; f(\mathbf{x}) \triangleq -x_1$$

$$\text{s.t. } g_1(\mathbf{x}) \triangleq -(1 - x_1)^3 + x_2 \le 0$$

$$g_2(\mathbf{x}) \triangleq -x_2 \le 0$$



What's going on here?

FONC for Equality & Inequality Constraints

- Let $f, g_1, \dots, g_{m_\text{i}}, h_1, \dots, h_{m_\text{e}}$ be $\mathcal{C}^1$
- Suppose that $\mathbf{x}^*$ is a (local) optimum point for

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize:}} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1, \dots, m_\text{i} \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, m_\text{e} \end{aligned}$$

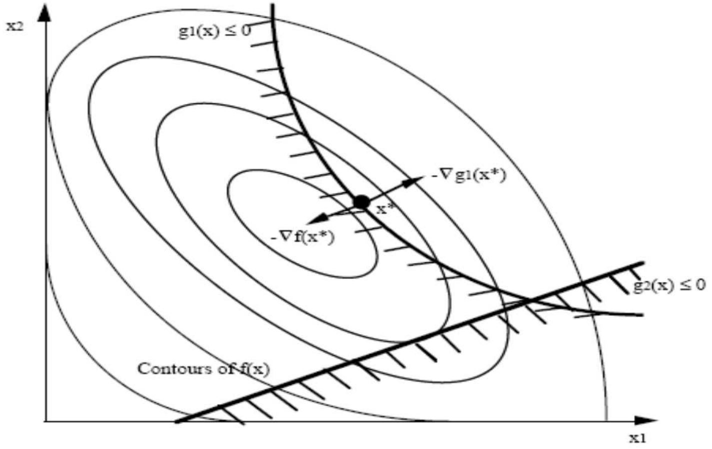
  and a regular point for the equality and active inequality constraints
- Then, there exist (unique) multiplier vectors $\boldsymbol{\nu}^* \in \mathbb{R}^{m_\text{i}}$, $\boldsymbol{\lambda}^* \in \mathbb{R}^{m_\text{e}}$ such that $(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$ satisfy:

$$\begin{array}{c} g(\mathbf{x}^*) \le 0, \quad \boldsymbol{\nu}^* \le 0, \quad \nu_i^* g_i(\mathbf{x}^*) = 0, \; i = 1, \dots, m_\text{i} \\ \mathbf{h}(\mathbf{x}^*) = 0, \quad \nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = 0 \end{array}$$

where $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) - \sum_{i=1}^{m_\text{i}} \nu_i^* g_i(\mathbf{x}^*) - \sum_{i=1}^{m_\text{e}} \lambda_i^* h_i(\mathbf{x}^*)$.

Understanding Necessary Conditions for Optimality

$$\min_{\mathbf{x}} \; f(\mathbf{x})$$
$$\text{s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$

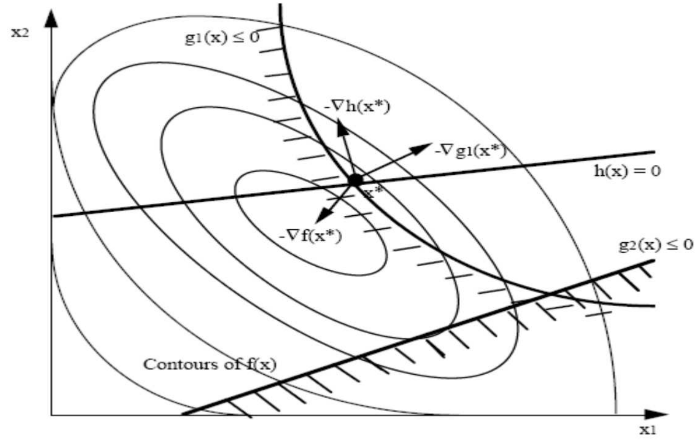


Analogy

- *"Ball rolling down valley pinned by fence"*
  - ➤ Balance of forces $(\nabla f, \nabla g_1)$

Understanding Necessary Conditions for Optimality

$$\min_{\mathbf{x}} \; f(\mathbf{x})$$
$$\text{s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$$
$$h(\mathbf{x}) = \mathbf{0}$$



Analogy

- *"Ball rolling down valley on rail pinned by fence"*
  - ➤ Balance of forces $(\nabla f, \nabla g_1, \nabla h)$

Second-Order Sufficient Conditions for a Local Minimum

- Let $f, g_1, \ldots, g_{m_i}, h_1, \ldots, h_{m_e}$ be $\mathcal{C}^2$
- Suppose that $\mathbf{x}^*$ is regular and $\exists\, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ such that $\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*$ is a KKT point
- For any $\mathbf{y} \neq 0$ such that

$$\nabla g_i(\mathbf{x}^*)^T \mathbf{y} = 0, \quad i \in \mathbb{A}\mathbf{x}^*, \quad \nu_i^* > 0,$$
$$\nabla g_i(\mathbf{x}^*)^T \mathbf{y} \leq 0, \quad i \in \mathbb{A}\mathbf{x}^*, \quad \nu_i^* = 0,$$
$$\nabla h(\mathbf{x}^*)^T \mathbf{y} = 0,$$

the inequality $\mathbf{y}^T \nabla^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)\mathbf{y} \succ 0$ holds.

Then, $\mathbf{x}^*$ is a local minimum.

More About Constraint Qualifications

Point $\mathbf{x}^*$ is regular if it satisfies some assumptions – Among the most used ones:

- **Linear Independence Constraint Qualification (LICQ)** The gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at $\mathbf{x}^*$,
  $$\text{rank}\left\{\nabla h_i(\bar{\mathbf{x}}), i = 1, \ldots, m_e, \nabla g_j(\bar{\mathbf{x}}), j \in \mathcal{A}\mathbf{x}^*\right\} = m_e + |\mathcal{A}\mathbf{x}^*|$$
- **Mangasarian-Fromovitz Constraint Qualification (MFCQ)** The gradients of the active inequality constraints and the gradients of the equality constraints are positive-linearly independent at $\mathbf{x}^*$,
  $$\exists \mathbf{d} \in \mathbb{R}^{n_x} \text{ such that } \nabla \mathbf{h}(\mathbf{x}^*)^{\mathsf{T}} \mathbf{d} = \mathbf{0} \text{ and } \nabla g_j(\mathbf{x}^*)^{\mathsf{T}} \mathbf{d} < 0, \forall j \in \mathcal{A}\mathbf{x}^*$$
- **Linearity Constraint Qualification** If g_i and h_i are affine functions, then no other CQ is needed
- **Slater Constraint Qualification** For a convex NLP only,
  $\exists \bar{\mathbf{x}}$ such that $\mathbf{h}(\bar{\mathbf{x}}) = 0$ and $g_j(\bar{\mathbf{x}}) < 0, \forall j \in \mathcal{A}\mathbf{x}^* \equiv$ "strict" feasibility !!

Convexity & First-Order Conditions for Optimality

- Consider the NLP problem

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize:}} \quad & f(\mathbf{x}) \\ \text{subject to:} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m_\text{i} \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m_\text{e} \end{aligned}$$

  where:
  - $f, g_1, \ldots, g_{m_\text{i}}, h_1, \ldots, h_{m_\text{e}}$ are $\mathcal{C}^1$
  - $f, g_1, \ldots, g_{m_\text{i}}$ are convex and $h_1, \ldots, h_{m_\text{e}}$ are affine on $\mathbb{R}^n$

- Suppose that $(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$ satisfy:

$$\boxed{\begin{array}{c} g(\mathbf{x}^*) \leq 0, \quad \boldsymbol{\nu}^* \leq 0, \quad \nu_i^* g_i(\mathbf{x}^*) = 0, \; i = 1, \ldots, m_\text{i} \\ \mathbf{h}(\mathbf{x}^*) = 0, \quad \nabla\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = 0 \end{array}}$$

  where $\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) - \sum_{i=1}^{m_\text{i}} \nu_i^* g_i(\mathbf{x}^*) - \sum_{i=1}^{m_\text{e}} \lambda_i^* h_i(\mathbf{x}^*)$.

Then, $\mathbf{x}^*$ is a global optimum point

Constraint qualification for convex problems ?!?

Sufficient condition for optimality

If $(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$ is a KKT point $\Rightarrow$ $\mathbf{x}^*$ is a global minimum

Why should we care about constraint qualification (CQ) ?!?

Necessary condition for optimality

If $\mathbf{x}^*$ is a minimum and satisfy (any) CQ

$\Rightarrow$

$\exists\, \lambda^*, \nu^*$ s.t. $(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$ is a KKT point

I.e. no CQ $\Rightarrow$ KKT may not have a solution even if $\mathbf{x}^*$ exists

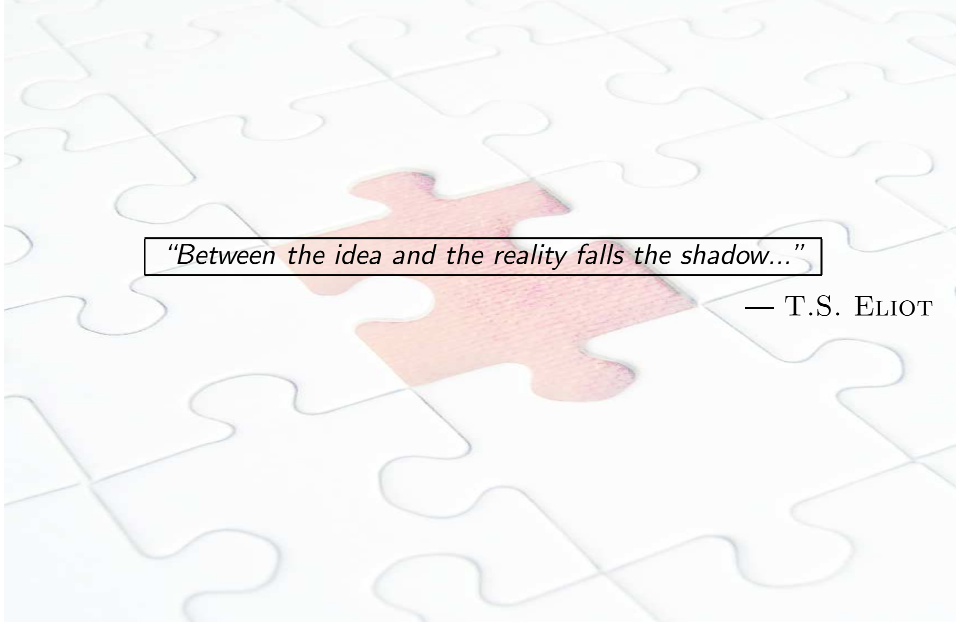
Summary of Optimality Conditions

Optimality conditions for NLP with equality and/or inequality constraints:

- **1st-Order Necessary Conditions**: A local optimum of a (differentiable) NLP must be a KKT point **if**:
  - that point is regular; **or**
  - all the (active) constraints are affine
- **1st-Order Sufficient Conditions**: A KKT point of a convex (differentiable) NLP is a global optimum
- **2nd-Order Necessary and Sufficient Conditions** requires positivity of the Hessian in the feasible directions

Non-convex problem $\Rightarrow$ minimum is not necessarily global.
But some non-convex problems have a unique minimum !!

Convex Optimization



How to Detect Convexity?

Reminder:

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \end{array}$$

is convex iff f, g_i are convex, h_i are affine.

Gradient and Hessian Tests

A function $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$ in $\mathcal{C}^2$ is convex on S if, at each $\mathbf{x}^\circ \in S$,

- **Gradient Test:** $f(\mathbf{x}) \geq f(\mathbf{x}^\circ) + \nabla f(\mathbf{x}^\circ)^\mathsf{T} (\mathbf{x} - \mathbf{x}^\circ), \quad \forall \mathbf{x} \in S$
- **Hessian Test:** $\nabla^2 f(\mathbf{x}^\circ) \succeq 0$ (positive semi-definite)

Strict convexity is detected by making the inequality signs strict

Gradient and Hessian tests are often very difficult !!

How to Detect Convexity?

Operations on sets preserving convexity

- Intersection $S_1 \cap S_2$ is convex is S_1, S_2 are convex
- $f(S) = \{f(x) | x \in S\}$ is convex if f is affine and S convex

Operations on functions preserving convexity

- Sum of functions: $f = \sum_i w_i f_i$ is convex if f_i are convex and $w_i \geq 0$ (extends to infinite sums & integrals)
- Affine composition: $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$ is convex if f is convex
- Point-wise maximum: $f(x) = \max_i \{f_i(x)\}$ is convex if f_i are convex
- Infimum: $g(x) = \inf_{y \in \mathcal{C}} f(x, y)$ is convex if $f(x, y)$ is convex in x, y
- Perspective function: $g(x, t) = tf(x/t)$ is convex on $t > 0$ if f is convex
- Composition: $g(f(x))$ is convex if f is convex and g is convex and monotonically increasing

Dual function

Consider the NLP:

$$\begin{array}{rl} f^* = \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \end{array}$$

with associated Lagrangian[2]: $\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\nu}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})$

The Lagrange dual function $g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \in \mathbb{R}$ is define as:

$$g(\boldsymbol{\nu}, \boldsymbol{\lambda}) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda})$$

- dual function $g(\boldsymbol{\nu}, \boldsymbol{\lambda})$ is concave (even if the NLP is non-convex !)
- can take values $-\infty$ (when Lagrangian unbounded below)
- lower bound $f^* \geq g(\boldsymbol{\nu}, \boldsymbol{\lambda})$, $\forall \boldsymbol{\nu} \geq 0$ and $\boldsymbol{\lambda}$

[2] signs in the Lagrange function are positive to abide by the literature

Dual problem

Primal problem:

$$\begin{array}{rl} f^* = \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \leq 0 \\ & \mathbf{h}(\mathbf{x}) = 0 \end{array}$$

with Lagrange dual function $g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \leq f^*$. What is the *best* lower bound ?

Lagrange dual problem:

$$\begin{array}{rl} g^* = \max_{\boldsymbol{\nu}, \boldsymbol{\lambda}} & g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \\ \text{s.t.} & \boldsymbol{\nu} \geq 0 \end{array}$$

- Dual problem is convex (even if primal problem is not !!)
- Weak duality $g^* \leq f^*$ always hold (even if primal problem is non-convex !!)
- Strong duality $g^* = f^*$ holds if primal is convex and some CQ holds
- Difference $f^* - g^* \geq 0$ is named *duality gap*

Saddle-point interpretation

Observe that:

$$\sup_{\nu\geq 0} \mathcal{L}(\mathbf{x},\boldsymbol{\nu}) = \sup_{\nu\geq 0}\left(f(\mathbf{x}) + \boldsymbol{\nu}^T\mathbf{g}(\mathbf{x})\right) = \begin{cases} f(\mathbf{x}) & \mathbf{g}(\mathbf{x}) \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

$$f^* = \inf_{\mathbf{x}} \sup_{\nu\geq 0} \mathcal{L}(\mathbf{x},\boldsymbol{\nu}) \underbrace{=}_{\text{strong duality}} g^* = \sup_{\nu\geq 0} g(\boldsymbol{\nu},\boldsymbol{\lambda}) = \sup_{\nu\geq 0}\inf_{\mathbf{x}} \mathcal{L}(\mathbf{x},\boldsymbol{\nu},\boldsymbol{\lambda})$$

As a result $\mathcal{L}(\mathbf{x},\boldsymbol{\nu})$ has the saddle-point property

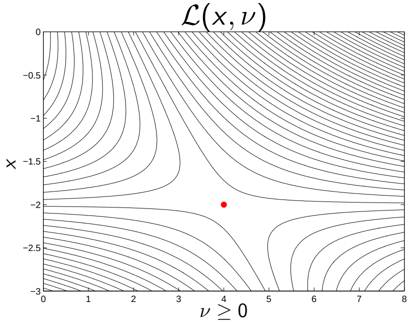
Example:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & x^2 \\ s.t. & x \leq -2 \end{array}$$

Lagrange function:

$$\mathcal{L}(x,\nu) = x^2 + \nu\,(x+2)$$

Optimum: $x^* = -2$, $\nu^* = 4$



Some typical Convex Problems

LP:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & A\mathbf{x} + \mathbf{b} = 0 \\ & C\mathbf{x} + \mathbf{d} \leq 0 \end{array}$$

QP:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \frac{1}{2}\mathbf{x}^T B\mathbf{x} + \mathbf{f}^T\mathbf{x} \\ \text{s.t.} & A\mathbf{x} + \mathbf{b} = 0 \\ & C\mathbf{x} + \mathbf{d} \leq 0 \end{array}$$

QCQP:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \frac{1}{2}\mathbf{x}^T B\mathbf{x} + \mathbf{f}^T\mathbf{x} \\ \text{s.t.} & A\mathbf{x} + \mathbf{b} = 0 \\ & \frac{1}{2}\mathbf{x}^T B_i\mathbf{x} + \mathbf{f}_i^T\mathbf{x} + d_i \leq 0 \end{array}$$

SOCP:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \mathbf{f}^T\mathbf{x} \\ \text{s.t.} & \|A_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T\mathbf{x} + d_i \\ & C\mathbf{x} + \mathbf{d} \leq 0 \end{array}$$

Inclusion: LP $\subset$ QP $\subset$ QCQP $\subset$ SOCP

Some more Convex Problems

Linear-fractional programming (e.g. *Von Neumann growth problem*)

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \dfrac{\mathbf{c}^T\mathbf{x} + d}{\mathbf{e}^T\mathbf{x} + f} \\ \text{s.t.} & G\mathbf{x} \leq 0 \end{array}$$

Geometric programming (convex after reformulation for $c_k > 0$)

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \sum_{k=1}^{K} c_k \prod_{i=1}^{n} x_i^{\alpha_{ik}} \\ \text{s.t.} & \sum_{k=1}^{K} c_k^j \prod_{i=1}^{n} x_i^{\alpha_{ik}^j} \leq 1, \quad j = 1,...,p \end{array}$$

Frobenius norm diagonal scaling: find D diagonal to minimize $\|DMD^{-1}\|_F^2$

$$\|DMD^{-1}\|_F^2 = \sum_{i,j} M_{ij}^2 d_i^2 d_j^{-2}$$

Generalized Inequalities

A cone $K \subseteq \mathbb{R}^n$ is proper if

- K is convex, closed and solid (i.e. nonempty interior) and
- K is pointed (i.e. if $x \in K$ and $-x \in K$ then $x = 0$)

A proper cone defines a partial ordering on $\mathbb{R}^n$ or generalized inequality:

$$x \leq_K y \Leftrightarrow y - x \in K$$

Examples:

- $K = \mathbb{R}^+$ yields the standard ordering, i.e. $x \leq_{\mathbb{R}^+} y \Leftrightarrow x \leq y$
- $K = \mathbb{S}_+^n$ yields the matrix inequality, i.e. $X \leq_{\mathbb{S}_+^n} Y \Leftrightarrow X \preceq Y$

Linear Matrix Inequalities & Semi-Definite Programming

A inequality of the form :
$$B_0 + \sum_{i=1}^{N} B_i x_i \succeq 0$$

with $B_i \in \mathbb{S}^n$, $i = 0, .., N$ is called an LMI

An optimization problem of the form:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & A\mathbf{x} - \mathbf{b} = 0 \\ & B_0 + \sum_{i=1}^{N} B_u x_i \succeq 0 \end{array}$$

is called an SDP.

It can be shown that LPs, QPs, QCQPs are SDPs !!

Example of SDPs - Eigenvalue problems

Consider the problem:

$$\min_{\mathbf{x} \in \mathcal{D}} \quad \lambda_{\max}\{G(\mathbf{x})\}$$

where $G(x) = B_0 + \sum_{i=1}^{N} B_i x_i$ and $B_i \in \mathbb{S}^n$.

The problem can be reformulated as the SDP:

$$\begin{array}{ll} \min\limits_{s \in \mathbf{R}, \mathbf{x} \in \mathcal{D}} & s \\ \text{s.t.} & sI - B_0 - \sum_{i=1}^{N} B_i x_i \succeq 0 \end{array}$$

Some more eigenvalue problems that have an SDP formulation

- Minimizing the spread of eigenvalues, i.e. $\lambda_{\max} - \lambda_{\min}$
- Minimizing the condition number of G, i.e. $\lambda_{\max}/\lambda_{\min}$
- Minimizing $\|\boldsymbol{\lambda}\|_1$

Problem with L_1 norm

"Best tractable approximation of L_0 norm", provides sparse residual $\mathbf{x}$

Smooth Reformulation

Non-smooth L_1 problem:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \sum_{i=1}^{N} \|x_i\|_1 \\ \text{s.t.} & A\mathbf{x} + \mathbf{b} = 0 \end{array}$$

$\rightarrow$

Smooth reformulation

$$\begin{array}{ll} \min\limits_{\mathbf{x},\mathbf{s}} & \sum_{i=1}^{N} s_i \\ \text{s.t.} & -s_i \leq x_i \leq s_i \\ & A\mathbf{x} + \mathbf{b} = 0 \end{array}$$

Example: L_1 fitting

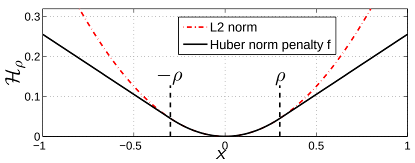
Find the best linear model $ax + b - y = 0$ for a set of data x_i,y_i

$$\begin{array}{ll} \min\limits_{a,b} & \|\mathbf{e}\|_1 \\ \text{s.t.} & a\mathbf{x} + b - \mathbf{y} = \mathbf{e} \end{array}$$

Problem with Huber Penalty function

Huber penalty function with $\rho > 0$:

$$\mathcal{H}_\rho(x) = \begin{cases} \frac{1}{2}x^2, & |x| \leq \rho \\ \rho(|z| - \frac{1}{2}\rho), & |x| > \rho \end{cases}$$



Estimation in the presence of outliers ("fat-tail" distribution)

Smooth Reformulation (Hessian not Lipschitz)

Non-smooth problem

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & \sum_{i=1}^{N} \mathcal{H}_\rho(x_i) \\ \text{s.t.} & A\mathbf{x} + \mathbf{b} = 0 \end{array}$$

$\rightarrow$

$$\begin{array}{ll} \min\limits_{\mathbf{x},\mathbf{s},\boldsymbol{\eta}} & \sum_{i=1}^{N} \frac{1}{2} s_i^2 + \rho \eta_i \\ \text{s.t.} & \eta_i \geq 0 \\ & -s_i - \eta_i \leq x_i \leq s_i + \eta_i \\ & A\mathbf{x} + \mathbf{b} = 0 \end{array}$$

Nuclear Norm

Min rank problem: $\boxed{\min_{X\in\mathcal{D}} \quad \mathrm{rank}\,(X)}$ with $X \in \mathbb{R}^{m\times n}$ is NP-hard.

Idea: approximate the rank by a Nuclear Norm: $\boxed{\min_{X\in\mathcal{D}} \quad \|X\|_*}$ where

$$\|X\|_* = \sum_{i=1}^{\min m,n} \sigma_i(X)$$

Properties of the Nuclear Norm

- Convex !!
- if $X \in \mathbb{S}_+$, then $\|X\|_* = \mathrm{Tr}\,(X)$
- For $X \in \mathbb{R}^{m\times n}$, $\|X\|_* \le t$ iff $\exists\, Y \in \mathbb{R}^{m\times m}$ and $Z \in \mathbb{R}^{n\times n}$ such that:

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \succeq 0, \quad \mathrm{Tr}\,(Y) + \mathrm{Tr}\,(Z) \le 2t$$

  i.e. nuclear norm minimization equivalent to LMI problems.
- $\mathrm{rank}\,(X) \ge \gamma^{-1}\|X\|_*$ on $\mathcal{D} = \{X \in \mathbb{R}^{m\times n} \,|\, \|X\| \le \gamma\}$, i.e. the Nuclear Norm problem provides a lower bound for the rank problem

Nuclear Norm - a basic example

Low-order identification of the FIR system:

$$y(t) = \sum_{\tau=t-r}^{t} h(t-\tau)u(\tau) + v(t)$$

Hankel matrix of the system ($n_H = r/2 \in \mathbb{N}$) yields $Y = H_h U$, with

$$H_h = \begin{bmatrix} h(0) & h(1) & \dots & h(r-n_H) \\ h(1) & h(2) & \dots & h(r-n_H+1) \\ \dots & \dots & \dots & \dots \\ h(n_H) & h(n_H+1) & \dots & h(r) \end{bmatrix}$$

Nuclear Norm formulation of the rank minimization:

$$\boxed{\begin{array}{ll} \min\limits_{h} & t \\ \text{s.t.} & \|H_h\|_* \le t \\ & \|Y^{\mathrm{meas}} - Y\|_F^2 \le \gamma \end{array}}$$

Recommended softwares

CVX (Stanford)

- Disciplined convex programming
- LP, QP, QCQP, QP, SDP, GP
- Exploits sparsity (to some extents)
- Exists in CodeGen version (http://cvxgen.com/)
- Matlab interface

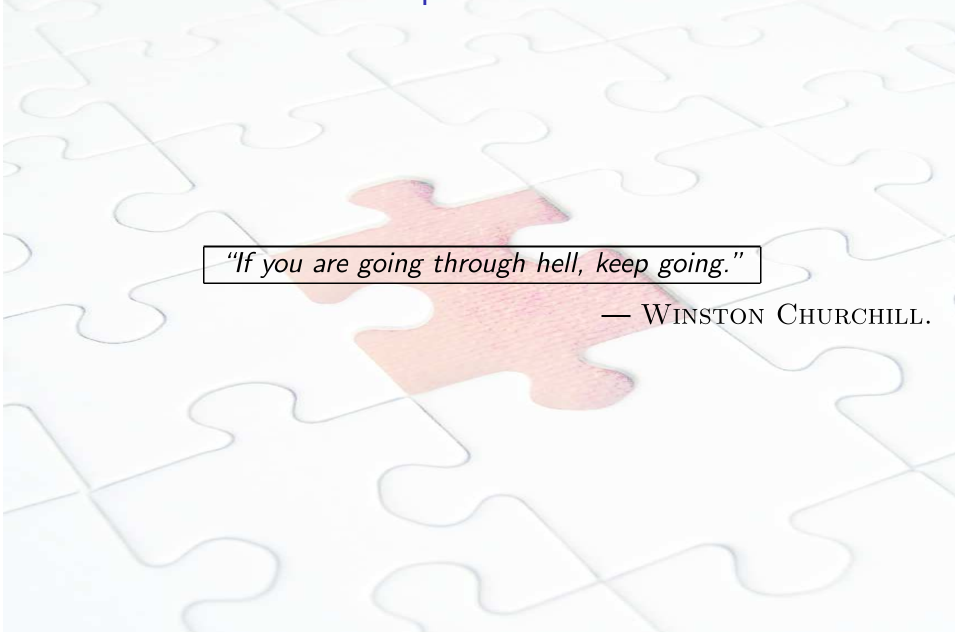
Sedumi (Matlab add-on)

- Optimization over symmetric cones
- Allows complex numbers
- Exploits sparsity for large-scale problems

Yalmip (Matlab add-on)

- Modeling interface for optimization
- Calls appropriate solvers

Numerical Methods for Optimization



Local optimization

A vast majority of solvers try to find an approximate KKT point...

Find the "primal-dual" variables $\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*$ such that:

$$\mathbf{g}(\mathbf{x}^*) \leq 0, \quad \boldsymbol{\nu}^* \leq 0, \quad \nu_i^* g_i(\mathbf{x}^*) = 0, \; i = 1, \ldots, m_{\text{i}}$$
$$\mathbf{h}(\mathbf{x}^*) = 0, \quad \nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = 0$$

Lets get started with the equality constrained problem

Find the "primal-dual" variables $\mathbf{x}^*, \boldsymbol{\lambda}^*$ such that:

$$\begin{aligned} \nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) &= 0, \\ \mathbf{h}(\mathbf{x}^*) &= 0. \end{aligned} \qquad (1)$$

Idea: apply a Newton search on the (non)linear system (1)

Newton method for equality constrained problems

The Newton recursion for solving the KKT conditions

$$\underbrace{\begin{bmatrix} \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) & \nabla \mathbf{h}(\mathbf{x}_k) \\ \nabla \mathbf{h}(\mathbf{x}_k)^T & 0 \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ -(\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k) \end{bmatrix} + \begin{bmatrix} \nabla \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \\ \mathbf{h}(\mathbf{x}_k) \end{bmatrix} = 0$$

Invertibility of the KKT matrix

The KKT matrix is invertible if (sufficient, not necessary)

- rank $\left(\nabla \mathbf{h}(\mathbf{x}_k)^T\right) = m$, with $\nabla \mathbf{h}(\mathbf{x}_k)^T \in \mathbb{R}^{m \times n}$ (LICQ)
- $\forall \mathbf{d} \neq 0$, such that $\nabla \mathbf{h}(\mathbf{x}_k)^T \mathbf{d} = 0$

$$\mathbf{d}^T \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \mathbf{d} \succ 0, \qquad \text{(SOSC)}$$

If $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is LICQ & SOSC, then the KKT matrix is invertible in a neighborhood of $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$

Newton method for equality constrained problems (cont')

Update of the dual variables

Define $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \Delta \boldsymbol{\lambda}_k$, and $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta \mathbf{x}_k$, observe that:

$$\nabla \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) = \nabla f(\mathbf{x}_k) - \nabla h(\mathbf{x}_k) \boldsymbol{\lambda}_k,$$

and use it in the KKT system

$$\begin{bmatrix} \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) & \nabla \mathbf{h}(\mathbf{x}_k) \\ \nabla \mathbf{h}(\mathbf{x}_k)^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k+1} - \mathbf{x}_k \\ -(\boldsymbol{\lambda}_{k+1} - \boldsymbol{\lambda}_k) \end{bmatrix} + \begin{bmatrix} \nabla \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \\ \mathbf{h}(\mathbf{x}_k) \end{bmatrix} = 0$$

KKT system in a "full dual update" form

$$\begin{bmatrix} \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) & \nabla \mathbf{h}(\mathbf{x}_k) \\ \nabla \mathbf{h}(\mathbf{x}_k)^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_k \\ -\boldsymbol{\lambda}_{k+1} \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}_k) \\ \mathbf{h}(\mathbf{x}_k) \end{bmatrix} = 0$$

The primal-dual iterate depends on $\boldsymbol{\lambda}_k$ only via the Hessian !!

Quadratic model interpretation

KKT system is a Quadratic Program (QP)

The iterate $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}$ is given by:

$$\min_{\mathbf{p} \in \mathbb{R}^n} \quad \frac{1}{2} \mathbf{d}^T \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \mathbf{d} + \nabla f(\mathbf{x}_k)^T \mathbf{d}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{x}_k) + \nabla \mathbf{h}(\mathbf{x}_k)^T \mathbf{d} = 0$$

Proof: KKT of the QP are equivalent to the KKT system.

Dual variables

Variables $\boldsymbol{\lambda}_{k+1}$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_{\text{QP}}$

Will be very usefull to tackle problems with inequality constraints !!

Failure of the full Newton step

Newton step $\Delta \mathbf{x}_k$ minimizes the quadratic model.

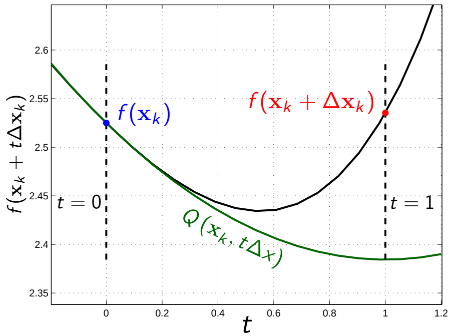
$$Q(\mathbf{x}_k, \Delta \mathbf{x}_k) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T \Delta \mathbf{x}_k + \frac{1}{2} \Delta \mathbf{x}_k^T \nabla^2 f(\mathbf{x}_k) \Delta \mathbf{x}_k$$

What if that model is not good enough ?

If Hessian $\nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \succ 0$ then

$$\nabla f(\mathbf{x}_k)^T \Delta \mathbf{x}_k = - \nabla f(\mathbf{x}_k)^T \nabla^2 \mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \nabla f(\mathbf{x}_k) < 0$$

i.e. the Newton step is a descent direction, but the full Newton step can increase the cost !!



A situation with $f(\mathbf{x}_k + \Delta \mathbf{x}_k) > f(\mathbf{x}_k)$ can easily occur...

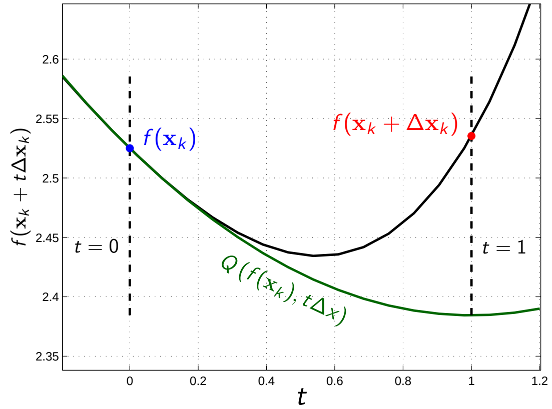
- Strong variation of $\nabla^2 f(\mathbf{x})$
- Nonlinear constraints

Globalization - Line search strategies

Exact line search (for unconstrained optimization)

Find the best step length:

$$t = \arg \min_{s \in]0,1]} f(\mathbf{x}_k + s\Delta \mathbf{x}_k)$$

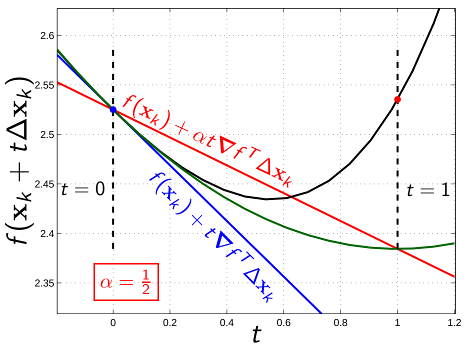


Globalization - Line search strategies

"Armijo's" backtracking line search (for unconstrained optimization)

Given a primal direction $\Delta \mathbf{x}_k$, using $0 < \alpha \leq \frac{1}{2}$ and $0 < \beta < 1$, do $t = 1$:

While: $f(\mathbf{x}_k + t\Delta \mathbf{x}_k) < f(\mathbf{x}_k) + \alpha t \nabla f(\mathbf{x}_k)^T \Delta \mathbf{x}_k$, **do:** $t = \beta t$



- If α too small we may accept steps yielding only mediocre improvement.
- If f quadratic, we want full step, i.e.

$$\alpha \leq \frac{1}{2}$$

Convergence of the Newton with Line-search (I)

Theorem

Assume that for $x, y \in S$:

- Hessian satisfies $m\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}$,
- and is Lipschitz, i.e. $\|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{y})\|_2 \leq L\|x - y\|_2$

then $\exists\ \eta, \gamma > 0$ with $\eta < \frac{m^2}{L}$ such that $\forall\ \mathbf{x}_k \in S$:

Damped phase:

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \leq -\gamma \quad \text{if} \quad \|\nabla f(\mathbf{x}_k)\|_2 \geq \eta$$

Quadratic phase:

$$\frac{L}{2m^2}\|\nabla f(\mathbf{x}_{k+1})\|_2 \leq \left(\frac{L}{2m^2}\|\nabla f(\mathbf{x}_{k+1})\|_2\right)^2 \quad \text{if} \quad \|\nabla f(\mathbf{x}_k)\|_2 < \eta$$

Two-phase convergence

- If $\mathbf{x}_k \in S$ is *far* from $\mathbf{x}^*$ $\Rightarrow$ Damped convergence (reduced steps)
- If $\mathbf{x}_k \in S$ is *close* to $\mathbf{x}^*$ $\Rightarrow$ Quadratic convergence (full steps)
- Once Newton has entered the quadratic phase, it stays quadratic !!

Affine invariance of the exact Newton method

Affine change of coordinates

Consider: $\mathbf{x} = T\mathbf{y} + \mathbf{t}$ with $T \in \mathbb{R}^{n\times n}$ non-singular and $\mathbf{t} \in \mathbb{R}^n$.
Define $\tilde{f}(\mathbf{y}) = f(T\mathbf{y}+\mathbf{t})$ and $\tilde{\mathbf{h}}(\mathbf{y}) = \mathbf{h}(T\mathbf{y}+\mathbf{t})$, then:

$$\nabla_{\mathbf{y}}\tilde{\mathcal{L}}(\mathbf{y},\boldsymbol{\lambda}) = T^T\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \quad \text{and} \quad \nabla^2_{\mathbf{yy}}\tilde{\mathcal{L}}(\mathbf{y},\boldsymbol{\lambda}) = T^T\nabla^2\mathcal{L}_{\mathbf{xx}}(\mathbf{x},\boldsymbol{\lambda})T$$

It can be verified that:

$$\begin{bmatrix} \nabla^2_{\mathbf{yy}}\tilde{\mathcal{L}}(\mathbf{y}_k,\boldsymbol{\lambda}_k) & \nabla\tilde{\mathbf{h}}(\mathbf{y}_k) \\ \nabla\tilde{\mathbf{h}}(\mathbf{y}_k)^T & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y}_k \\ -\Delta\boldsymbol{\lambda}_k \end{bmatrix} + \begin{bmatrix} \nabla\tilde{\mathcal{L}}(\mathbf{y}_k,\boldsymbol{\lambda}_k) \\ \tilde{\mathbf{h}}(\mathbf{y}_k) \end{bmatrix} = 0$$

holds for $\Delta\mathbf{x}_k = T\Delta\mathbf{y}_k$.

The Newton step is invariant w.r.t. an affine change of coordinate.

Convergence of the exact Newton method (II)

Self-concordant functions

- Function $f \in \mathbb{R}$ convex is self-concordant iff $\left|f^{(3)}(\mathbf{x})\right| \leq 2f^{(2)}(\mathbf{x})^{3/2}$.
- Function $\mathbf{f} \in \mathbb{R}^n$ convex is self-concordant iff $\tilde{f}(t) = \mathbf{f}(\mathbf{x}+t\mathbf{v})$ is self-concordant for all $\mathbf{x} \in \text{Dom}(\mathbf{f})$ and $\mathbf{v} \in \mathbb{R}^n$.

Self-concordance theory

Define $\xi = \left(\nabla f(\mathbf{x})^T\nabla^2 f(\mathbf{x})^{-1}\nabla f(\mathbf{x})\right)^{\frac{1}{2}}$.

Starting from $\mathbf{x}_0$, assume that:

- f is strictly convex, sublevel set $S = \{\mathbf{x}|f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ is closed.

Then $\exists\,\eta,\gamma > 0$ with $0 < \eta \leq \frac{1}{4}$ s.t.[a]

- If $\xi(\mathbf{x}_k) > \eta$, then $f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \leq -\gamma$ (damped phase)
- If $\xi(\mathbf{x}_k) > \eta$, then $2\xi(\mathbf{x}_{k+1}) \leq (2\xi(\mathbf{x}_k))^2$ (quadratic phase)

[a] η,γ depend only on the line search parameters)

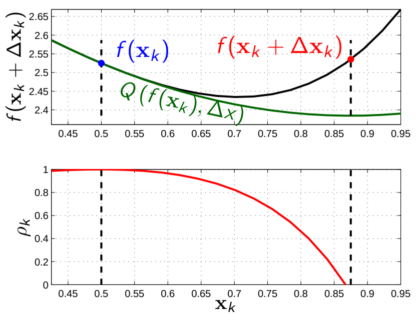
Globalization - Trust-region methods

"Trustworthiness" of the quadratic model

$$\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k+\Delta\mathbf{x}_k)}{Q(\mathbf{x}_k) - Q(\mathbf{x}_k+\Delta\mathbf{x}_k)}, \qquad \begin{array}{ll} \rho_k = 1, & \text{perfect model} \\ \rho_k < 0, & f(\mathbf{x}_k+\Delta\mathbf{x}_k) > f(\mathbf{x}_k) \end{array}$$

Idea: adjust the direction with the step length

Illustrative example with $\mathbf{x} \in \mathbb{R}$



Trust-region solves:

$$\Delta\mathbf{x}_k = \arg\min_p \; Q(\mathbf{x}_k,p) \quad \text{s.t.} \quad \|p\| \leq \Delta_k$$

- Line-search: get the direction, decide the length
- Trust-region: decide the length, find the direction

Globalization - Trust-region methods

$$\Delta\mathbf{x}_k = \arg\min_p \; Q(\mathbf{x}_k,p) \quad \text{s.t.} \quad \|p\| \leq \Delta_k \qquad (2)$$

$$\rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_k+\Delta\mathbf{x}_k)}{Q(\mathbf{x}_k) - Q(\mathbf{x}_k+\Delta\mathbf{x}_k)}$$

Trust-region Algorithm - Heuristic to choose Δ_k from observed ρ_k

Inputs: $\Delta_{\max}$, $\eta \in [0, 0.25]$, Δ_0, $\mathbf{x}_0$, TOL > 0

while $\|\nabla f(\mathbf{x}_k)\| > \text{TOL}$, **do**:

Get $\Delta\mathbf{x}_k$ from (2)

Evaluate $f(\mathbf{x}_k+\Delta\mathbf{x}_k)$, compute ρ_k

Length adaptation: $\Delta_{k+1} = \begin{cases} 0.25\Delta_k & \text{if} \quad \rho_k < 0.25 \\ \min(2\Delta_k,\Delta_{\max}) & \text{if} \quad \rho_k > 0.75 \\ \Delta_k & \text{if} \quad \text{otherwise} \end{cases}$

Decide acceptance: $\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_{k+1}+\Delta\mathbf{x}_k & \text{if} \quad \rho_k > \eta \\ \mathbf{x}_{k+1} & \text{if} \quad \rho_k \leq \eta \end{cases}$

$k = k+1$

end while

Newton-type Methods

Computing $\nabla^2\mathcal{L}$ is expensive, use an approximation B_k instead !!

Descent direction

If $B_k \succ 0$ then $\Delta \mathbf{x}_k = -B_k^{-1}\nabla f(\mathbf{x}_k)$ is a descent direction.

Local convergence for Newton-type Methods

Assume

- $\mathbf{x}^*$ is SOSC for f
- Lipschitz condition: $\|B_k^{-1}\left(\nabla^2 f(\mathbf{x}_k) - \nabla^2 f(\mathbf{y})\right)\| \leq \omega\|\mathbf{x}_k - \mathbf{y}\|$ holds on the sequence $k = 0, 1, ..., y \in \mathbb{R}^n$.
- Compatibility: $\|B_k^{-1}\left(\nabla^2 f(\mathbf{x}_k) - B_k\right)\| \leq \kappa_k$ with $\kappa_k \leq \kappa < 1$

Then if $\mathbf{x}_k$ is close to $\mathbf{x}^*$, $\mathbf{x}_k \to \mathbf{x}^*$ and convergence is

- Quadratic for $\kappa = 0$: $\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq C\|\mathbf{x}_k - \mathbf{x}^*\|^2$ with $C = \omega/2$
- Superlinear for $\kappa_k \to 0$: $\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq C_k\|\mathbf{x}_k - \mathbf{x}^*\|$ with $C_k \to 0$
- Linear for $\kappa_k > \rho$: $\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq C\|\mathbf{x}_k - \mathbf{x}^*\|$ with $C < 1$

Steepest descent

Constant Hessian approximation

Use $B_k = \alpha_k^{-1} I$, then:

$$\Delta \mathbf{x}_k = -B_k^{-1}\nabla f(\mathbf{x}_k) = -\alpha_k \nabla f(\mathbf{x}_k)$$

Step size α_k is chosen sufficiently small by the line-search.

Convergence

- Compatibility: $\|\alpha_k\left(\nabla^2 f(\mathbf{x}_k) - I\right)\| \leq \kappa_k$ with $\kappa_k \leq \kappa < 1$
- Constant does not converge to 0, i.e. $\kappa_k > \rho, \ \forall k$
- Linear convergence when $\mathbf{x}_k$ is close to $\mathbf{x}^*$

Gauss-Newton Method

Cost function of the type $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{F}(\mathbf{x})\|_2^2$, with $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2 f(\mathbf{x}) = \frac{\partial}{\partial x}\left(\nabla F(\mathbf{x}) F(\mathbf{x})\right) = \nabla F(\mathbf{x})\nabla F(\mathbf{x})^T + \sum_{i=1}^{m} \nabla^2 F_i(\mathbf{x}) F_i(\mathbf{x})$$

Gauss-Newton method proposes to use: $B_k = \nabla F(\mathbf{x}_k)\nabla F(\mathbf{x}_k)^T + \alpha_k I$

B_k is a good approximation if:

- all $\nabla^2 F_i(\mathbf{x})$ are small (F close to linear), or
- all $F_i(\mathbf{x})$ are small

Typical application to fitting problems: $F(\mathbf{x}) = \sum_{i=1}^{N} \|\mathbf{y}_i(\mathbf{x}) - \bar{\mathbf{y}}_i\|_2^2$

Convergence

- If $\sum_{i=1}^{m} \nabla^2 F_i(\mathbf{x}) F_i(\mathbf{x}) \to 0$ then $\kappa_k \to 0$
- Quadratic convergence when $\mathbf{x}_k$ is close to $\mathbf{x}^*$

Quasi-Newton Methods

Compute numerical derivative of $\nabla^2 f(\mathbf{x})$ in an efficient (iterative) way

BFGS

Define

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$
$$\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$$

Idea: Update $B_k \to B_{k+1}$ such that $B_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula[a]: $B_{k+1} = B_k - \dfrac{B_k \mathbf{s}\mathbf{s}^T B_k}{\mathbf{s}^T B_k \mathbf{s}} + \dfrac{\mathbf{y}\mathbf{y}^T}{\mathbf{s}^T\mathbf{y}}, \quad B_0 = I$

[a] See "Powell's trick" to make sure that $B_{k+1} > 0$

Convergence

- It can be shown that $B_k \to \nabla^2 f(\mathbf{x})$, then $\kappa_k \to 0$
- Quadratic convergence when $\mathbf{x}_k$ is close to $\mathbf{x}^*$

What about inequality constraints ?

Find the "primal-dual" variables $\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*$ such that:

$$g(\mathbf{x}^*) \leq 0, \quad \boldsymbol{\nu}^* \leq 0, \quad \nu_i^* g_i(\mathbf{x}^*) = 0, \\ \mathbf{h}(\mathbf{x}^*) = 0, \quad \nabla\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = 0$$

Conditions $\nu_i^* g_i(\mathbf{x}^*) = 0$ are not smooth !!

Active set methods - Outline of the idea

Guess the active set $\mathbb{A}\mathbf{x}^*$ a priori,

Solve :

$$g_i(\mathbf{x}^*) = 0, i \in \mathbb{A} \\ \mathbf{h}(\mathbf{x}^*) = 0, \quad \nabla\mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) = 0$$

Check : $\boldsymbol{\nu}^* \leq 0$, and $g_i(\mathbf{x}^*) \leq 0, i \in \mathbb{A}^c$

If fails : adapt $\mathbb{A}$, back to solve.

Efficient only for Quadratic Programs !!

Quadratic Programming via Active Set Method

$$\min_{\mathbf{x}} \quad \frac{1}{2}\mathbf{x}^T B\mathbf{x} + \mathbf{f}^T\mathbf{x} \\ \text{s.t.}: \quad A\mathbf{x} + \mathbf{b} \leq 0 \\ C\mathbf{x} + \mathbf{d} = 0$$

Active set methods for QP

Guess the active set $\mathbb{A}\mathbf{x}^*$ a priori,

Solve :

$$\begin{aligned} B\mathbf{x} + \mathbf{f} - A^T\boldsymbol{\nu} - C^T\boldsymbol{\lambda} &= 0 \\ A\mathbf{x} + \mathbf{b} &= 0 \\ C\mathbf{x} + \mathbf{d} &= 0 \end{aligned}$$

Check : $\boldsymbol{\nu} \leq 0$, and $g_i(\mathbf{x}) \leq 0, i \in \mathbb{A}^c$

If fails : adapt[a] $\mathbb{A}$, back to solve.

[a] many different techniques

Each iteration requires *only* to perform some linear algebra...

QP via the Primal Active Set Method

Primal Active Set Algorithm (e.g. qpOASES)

Starting from $\mathbf{x}_0$, active set $\mathbb{A}_0$, $k = 0$:

while $\mathbf{x}_k$ *is not a solution*, **do**:

Solve: $B\tilde{\mathbf{x}} + \mathbf{d} - A_{\mathbb{A}_k}^T\tilde{\boldsymbol{\nu}} = 0, \quad A\tilde{\mathbf{x}} + \mathbf{b}_{\mathbb{A}_k} = 0$

Find the max $t \in [0, 1]$ such that $\mathbf{x}_{k+1} = \mathbf{x}_k + t\,(\tilde{\mathbf{x}} - \mathbf{x}_k)$ is feasible

If $t < 1 \Rightarrow \mathbb{A}_{k+1} = \mathbb{A}_k \cup i_{\text{block}}$ (add blocking constraint)

If $t = 1$ ($\tilde{\mathbf{x}}$ primal feasible)

If $\tilde{\boldsymbol{\nu}} \leq 0 \Rightarrow$ solution found, exit algorithm

If $\tilde{\nu}_i > 0 \Rightarrow \mathbb{A}_{k+1} = \mathbb{A}_k - \{i\}$ (remove blocking constraints)

k = k+1

end while

Sequential Quadratic Programming

Consider the NLP...

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \\ \text{s.t} \quad \mathbf{g}(\mathbf{x}) \leq 0 \\ \mathbf{h}(\mathbf{x}) = 0$$

...and the corresponding QP

$$\min_{\mathbf{d}} \quad \frac{1}{2}\mathbf{d}^T B_k\mathbf{d} + \nabla f(\mathbf{x}_k)^T\mathbf{d} \\ \text{s.t} \quad \mathbf{g}(\mathbf{x}_k) + \nabla\mathbf{g}(\mathbf{x}_k)^T\mathbf{d} \leq 0 \\ \mathbf{h}(\mathbf{x}_k) + \nabla\mathbf{h}(\mathbf{x}_k)^T\mathbf{d} = 0$$

Theorem

Suppose

- The solution of the NLP $\mathbf{x}^*$ with active set $\mathbb{A}^*$ is LICQ,
- $\boldsymbol{\nu}^*$, $\mathbf{g}$ have strict complementarity,
- $\mathbf{x}_k$ is close enough to $\mathbf{x}^*$,
- $B_k \succeq 0$, and $B_k \succ 0$ on the nullspace of $\nabla g_{\mathbb{A}^*}$

then the QP has the active set $\mathbb{A}^*$ and strict complementarity.

Sequential Quadratic Programming

Monitoring progress with the L_1 merit function:

$$T_1(\mathbf{x}_k) = f(\mathbf{x}_k) + \mu\|\mathbf{h}(\mathbf{x}_k)\|_1 + \mu\sum_{i=1}^m |\min(0, g_i(\mathbf{x_k})|$$

Line-search SQP algorithm

while $T_1(\mathbf{x}_k) > \text{TOL}$ **do**
- get $\nabla f(\mathbf{x}_k)$, $\nabla \mathbf{g}(\mathbf{x}_k)$, $B_k \approx \nabla^2\mathcal{L}(\mathbf{x}_k, \boldsymbol{\nu}_k, \boldsymbol{\lambda}_k)$
- solve the QP, get $\mathbf{d}$, $\boldsymbol{\lambda}_{\text{QP}}$, $\boldsymbol{\nu}_{\text{QP}}$
- perform line-search on $T_1(\mathbf{x}_k + \mathbf{d})$, get step length α
- take primal step: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\mathbf{d}$
- take dual step: $\boldsymbol{\lambda}_{k+1} = (1-\alpha)\boldsymbol{\lambda}_k + \alpha\boldsymbol{\lambda}_{QP}$, $\boldsymbol{\nu}_{k+1} = (1-\alpha)\boldsymbol{\nu}_k + \alpha\boldsymbol{\nu}_{QP}$

end while

Theorem

If $\nabla^2\mathcal{L}(\mathbf{x}_k, \boldsymbol{\nu}_k, \boldsymbol{\lambda}_k) \succ 0$ and $\mu > \max\{\|\boldsymbol{\nu}_{k+1}\|_\infty, \|\boldsymbol{\lambda}_{k+1}\|_\infty\}$ then $\mathbf{d}$ is a descent direction for $T_1(\mathbf{x}_k)$

Primal-dual Interior Point Methods

Barrier method: introduce the inequality constraints in the cost function

Primal Interior point method

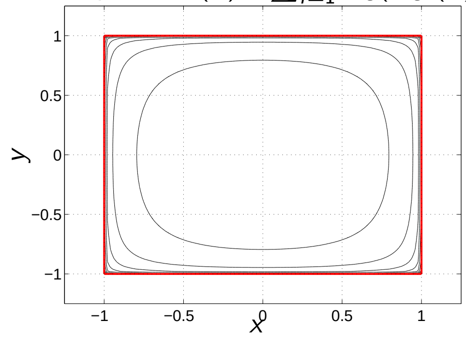
$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{g}(\mathbf{x}) \le 0 \end{array} \quad \rightarrow \quad \min_{\mathbf{x}_\tau} \; f(\mathbf{x}_\tau) - \tau\sum_{i=1}^{m_i}\log(-g_i(\mathbf{x}_\tau))$$

Example: box constraint

$$-1 \le x \le 1$$
$$-1 \le y \le 1$$

$$\lim_{\tau\to 0}\mathbf{x}_\tau^* = \mathbf{x}^*$$

Contour plot of $f(\mathbf{x}) - \sum_{i=1}^{m_i}\log(-g_i(\mathbf{x}))$



Primal-dual Interior Point Methods

KKT interpretation

KKT of the original problem

$$\nabla f(\mathbf{x}^*) - \nabla \mathbf{g}(\mathbf{x}^*)\boldsymbol{\nu}^* = 0$$
$$\mathbf{g}(\mathbf{x}^*) \le 0, \quad \boldsymbol{\nu}^* \le 0, \quad \nu_i^* g_i(\mathbf{x}^*) = 0$$

KKT of the barrier problem $\rightarrow \tilde{\nu}_i^* = \tau g_i(\mathbf{x}_\tau^*)^{-1} \rightarrow$ IP-KKT (primal-dual)

$$\nabla f(\mathbf{x}_\tau^*) - \tau\sum_{i=1}^{m_i}\nabla \mathbf{g}(\mathbf{x}_\tau^*)g_i(\mathbf{x}_\tau^*)^{-1} = 0 \qquad \begin{aligned} \nabla f(\mathbf{x}_\tau^*) - \nabla \mathbf{g}(\mathbf{x}_\tau^*)\tilde{\boldsymbol{\nu}}^* &= 0 \\ \tilde{\nu}_i^* g_i(\mathbf{x}_\tau^*) &= \tau \end{aligned}$$

A basic primal-dual IP algorithm

From $\mathbf{x}_0$, $\tau > 0$ *sufficiently* large

while *"stopping test for the original problem fails"* **do:**
- solve IP-KKT to TOL, $\mathbf{x}_{k+1} = \mathbf{x}_\tau^*$
- $\tau = \sigma\tau$, $\sigma \in]0, 1[$

end while

Primal-dual Interior Point Methods

Slack formulation - ensuring feasibility

Slack formulation of the IP-KKT:

$$\begin{aligned} \nabla f(\mathbf{x}_\tau^*) - \nabla \mathbf{g}(\mathbf{x}_\tau^*)\tilde{\boldsymbol{\nu}}^* &= 0 \\ \tilde{\nu}_i^* s_i^* &= \tau \\ \mathbf{g}(\mathbf{x}^*) - \mathbf{s}^* &= 0 \end{aligned}$$

Newton system (symmetrized):

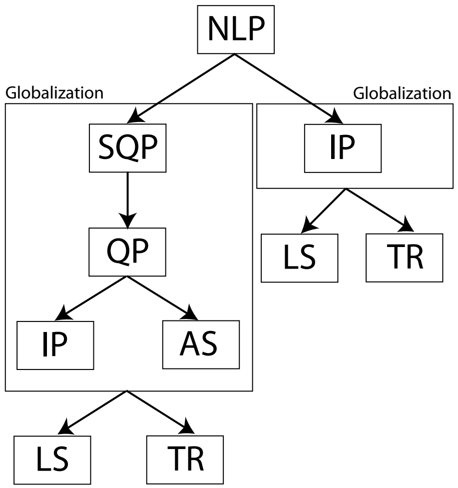
$$\begin{bmatrix} \nabla^2\mathcal{L} & 0 & \nabla g \\ 0 & \Sigma & -I \\ \nabla g & -I & 0 \end{bmatrix}\begin{bmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{s} \\ -\Delta\tilde{\boldsymbol{\nu}} \end{bmatrix} = -\begin{bmatrix} \nabla\mathcal{L} \\ \tilde{\boldsymbol{\nu}} - S^{-1}\tau \\ \mathbf{g}(\mathbf{x}) - \mathbf{s} \end{bmatrix}$$

where $\Sigma = \text{diag}\left(\tilde{\nu}_i s_i^{-1}\right)$ and $S = \text{diag}(\mathbf{s})$.

Sketch of the primal-dual IP algorithm

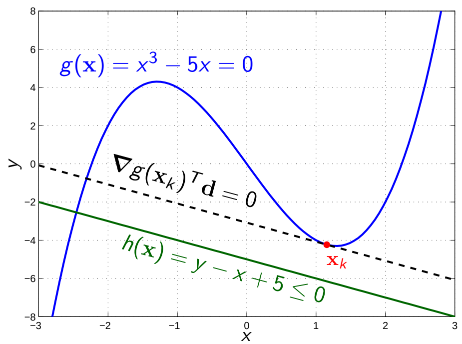
- Start with feasible guess $\mathbf{s} > 0$, $\tilde{\boldsymbol{\nu}} > 0$
- Line-search, enforce: $\mathbf{s}_{k+1} \ge (1-\tau)\mathbf{s}_k$ and $\tilde{\boldsymbol{\nu}}_{k+1} \ge (1-\tau)\tilde{\boldsymbol{\nu}}_k$

Summary of numerical optimization



IP = Primal-dual Interior-point method
SQP = Sequential Quadratic Programming
QP = Quadratic Programming
AS = Primal Active Set Method
LS = Line-search Method
TR = Trust-region Method

Failure of the methods - Infeasible points



Vectors $\nabla \mathbf{g}(\mathbf{x}_k)$ and $\nabla \mathbf{h}(\mathbf{x}_k)$ are not Lin. Independent, i.e.

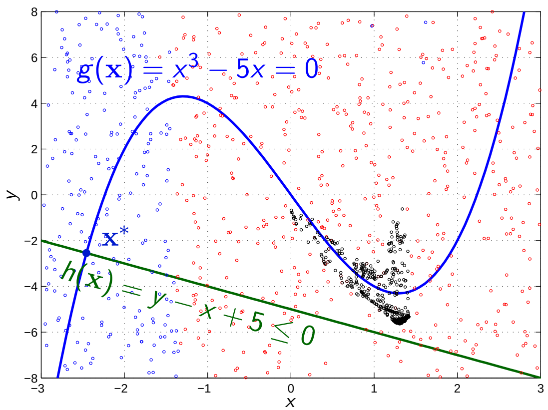
$\nexists\, \mathbf{d} \in \mathbb{R}^n$ such that:

$$\begin{bmatrix} \nabla \mathbf{g}(\mathbf{x}_k)^T \\ \nabla \mathbf{h}(\mathbf{x}_k)^T \end{bmatrix} \mathbf{d} + \begin{bmatrix} \mathbf{g}(\mathbf{x}_k) \\ \mathbf{h}(\mathbf{x}_k) \end{bmatrix} = 0$$

There is no feasible direction $\mathbf{d} \in \mathbb{R}^n$

Failure of the methods - Infeasible points

Interior-point method applied to the proposed example[3]



Problem:

$$\begin{aligned} \min_{x,y} \quad & x^2 + y^2 \\ \text{s.t.} \quad & x^3 - 5x = 0 \\ & y - x + 5 \leq 0 \end{aligned}$$

Red dots: failed starting points → black dots
Blue dots: succesfull starting points → $(-2.46, -2.54)$

[3] Similar results with SQP

Homotopy strategies

Parametric NLP

Consider the parametric NLP $P(\theta)$, $\theta \in \mathbb{R}$:

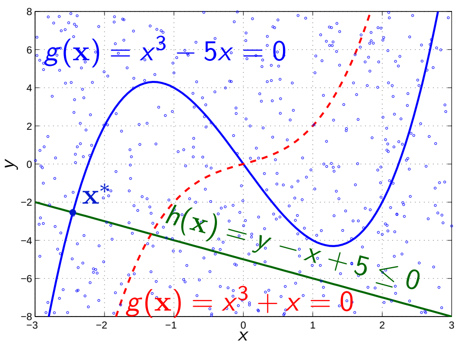
$$\begin{aligned} P(\theta) = \min_{\mathbf{x}} \quad & f(\mathbf{x}, \theta) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}, \theta) \leq 0 \\ & \mathbf{h}(\mathbf{x}, \theta) = 0 \end{aligned}$$

If $\nabla f, \nabla \mathbf{g}, \nabla \mathbf{h}$ are differentiable w.r.t. θ and the (parametric) solution $(\mathbf{x}^*(\theta), \boldsymbol{\nu}^*(\theta), \boldsymbol{\lambda}^*(\theta))$ is SOSC and LICQ, then it is differentiable w.r.t. θ.

Homotopy - Outline of the idea

Suppose that $P(1)$ is the NLP to be solved, $P(0)$ is an NLP that *can* be solved. Then starting from $\theta = 0$, solve $P(\theta)$ while gradually decreasing $\theta \rightarrow 1$. If LICQ & SOSC are maintained on the way, then a solution can be obtained.

Homotopy strategies



Note: SQP very efficient for homotopy strategies !!

Start with initial guess x_0, y_0
Set $\theta = 0$, $k = 0$
while $\theta < 1$ **do**:
Using x_k, y_k as initial guess solve:

$$\begin{aligned} P(\theta) = \min_{\mathbf{x}} \quad & x^2 + y^2 \\ \text{s.t.} \quad & x^3 - 5tx = 0 \\ & t = 1.2\theta - 0.2 \\ & y - x + 5 \leq 0 \end{aligned}$$

set $x_{k+1} = x^*(\theta_k), y_{k+1} = y^*(\theta_k)$
$\theta_{k+1} = \theta_k + 0.1$
$k = k + 1$
end while

Parameter embedding for homotopies

Consider running a homotopy on the parametric NLP, with $\boldsymbol{\theta} \in \mathbb{R}^p$:

$$\begin{aligned} P(\boldsymbol{\theta}) = \min_{\mathbf{x}} \quad & f(\mathbf{x}, \boldsymbol{\theta}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}, \boldsymbol{\theta}) \leq 0 \\ & \mathbf{h}(\mathbf{x}, \boldsymbol{\theta}) = 0 \end{aligned}$$

Then the parameters should be embedded in the NLP, i.e. solve

$$\begin{aligned} P^E(\boldsymbol{\theta}) = \min_{\mathbf{x}, \boldsymbol{\zeta}} \quad & f(\mathbf{x}, \boldsymbol{\zeta}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}, \boldsymbol{\zeta}) \leq 0 \\ & \mathbf{h}(\mathbf{x}, \boldsymbol{\zeta}) = 0 \\ & \boldsymbol{\zeta} - \boldsymbol{\theta} = 0 \end{aligned}$$

Because $\boldsymbol{\zeta}$ is part of the decision variables, the sensitivity of the cost and constraints w.r.t. $\boldsymbol{\zeta}$ is computed. That information is intrinsically used by the solver to update the solution $\mathbf{x}^*(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is changed.

Recommended softwares

ipopt

- Large-scale primal-dual IP solver, filter techniques, sparse linear algebra, extremely robust

KNITRO

- Large-scale solver
- Primal IP solver (direct factorization/CG) and Active-set solver
- Interface to many environments (AMPL, MATLAB, Mathematica,...)

SNOPT

- Large-scale SQP solver, augmented Lagrangian merit function

CasADi

- Symbolic framework for Automatic Differentiation (AD) tool
- Python interface, CAS syntax
- Interface to most state-of-the-art NLP solvers & intergators

Summary & additional remarks

- Beware of your constraints qualification !! Even a simple convex problem can fail state-of-the-art optimizers if badly formulated...
- Newton-type techniques converge really fast*... (*if started close to the solution).
- (Primal-dual) Interior-point methods: extensively used for both convex and non-convex problems, SDP, generalized constraints, well suited for large-scale problems
- Sequential-quadratic Programming: very powerful for parametric optimization problems, homotopies, optimal control
- Coming-back of (Parallel) First-order techniques for (very) large-scale problems
- Check-out existing (open-source) softwares before developing your own algorithm
- Strong non-convexity can often be overcome (homotopy strategies). Requires some insights in the problem though. Solution is local.

Some good readings

Convex Optimization, S. Boyd, L. Vandenberghe, Cambrige University Press
Nonlinear Programming, L.T. Biegler, MOS-SIAM
Numerical Optimization, T.V. Mikosch, S.I. Resnick, S.M. Robinson, Springer Series
Primal-Dual Interior-Point Methods, S.J. Wright, SIAM
Optimization Theory & Methods, W. Sun, Y. Yuan, Springer