Modern Optimization Techniques

4. Inequality Constrained Optimization / 4.2. Barrier and Penalty Methods

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute of Computer Science
University of Hildesheim, Germany

Syllabus

Mon. 30.10.	(0)	0. Overview
		1. Theory
Mon. 6.11.	(1)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Mon. 13.11.	(2)	2.1 Gradient Descent
Mon. 20.11.	(3)	2.2 Stochastic Gradient Descent
Mon. 27.11.	(4)	2.3 Newton's Method
Mon. 4.12.	(5)	2.4 Quasi-Newton Methods
Mon. 11.12.	(6)	2.5 Subgradient Methods
Mon. 18.12.	(7)	2.6 Coordinate Descent
	—	— *Christmas Break* —
		3. Equality Constrained Optimization
Mon. 8.1.	(8)	3.1 Duality
Mon. 15.1.	(9)	3.2 Methods
		4. Inequality Constrained Optimization
Mon. 22.1.	(10)	4.1 Primal Methods
Mon. 29.1.	(11)	4.2 Barrier and Penalty Methods
Mon. 5.2.	(12)	4.3 Cutting Plane Methods

Outline

Outline

Inequality Constrained Minimization (ICM) Problems

A problem of the form:

$$
\begin{aligned}
\arg\min_{x \in \mathbb{R}^N} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\
& h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q
\end{aligned}
$$

where:

- $f : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- $g_1, \ldots, g_P : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- $h_1, \ldots, h_Q : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Inequality Constrained Minimization (ICM) Problems / Affine Constraints

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} - a = 0 \\ & B\mathbf{x} - b \leq 0 \end{aligned}$$

where:

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $B \in \mathbb{R}^{Q \times N}, b \in \mathbb{R}^Q$: Q affine inequality constraints
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Barrier and Penalty Methods

- Barrier and Penalty methods reduce the problem to a
  - **sequence** of optimization problems
  - with a more complex objective function,
  - but with simpler constraints

- Applies a suitable optimization method to each of the problems
  - often Newton

Advantages:

1. Does not suffer from combinatorical complexity for many constraints (as primal methods / active set methods do)
2. Generally applicable, as they do not rely on special problem structure.

Outline

Idea

- search only in the interior of the feasible area S
  - ensure that an optimization algorithm stays within the interior by adding a barrier function B to the objective
    - the barrier B grows unbounded when approaching the border of the feasible area.
  - aka as interior point methods

- iteratively reduce the weight c of the barrier.
  - iterates $x^{(k)}$ converge to the optimum x^*, possibly on the border of the feasible area.

- only applicable if the interior of the feasible area is not empty, esp. there are no equality constraints.

Idea

For $f : S \to \mathbb{R}$ and $S \subseteq \mathbb{R}^N$:

$$x = \arg\min_{x \in S} f(\mathbf{x}) \quad \Longleftrightarrow \quad x = \lim x^{(k)}, \quad c^{(k)} \to 0$$

$$x^{(k)} := \arg\min_{x \in S^\circ} \tilde{f}_{c^{(k)}}(\mathbf{x})$$

$$\tilde{f}_c(x) := f(\mathbf{x}) + cB(\mathbf{x})$$

with a barrier function

$$B : S^\circ \to \mathbb{R}$$
$$(i) B \text{ continuous}$$
$$(ii) B(x) \geq 0$$
$$(iii) B(x) \to \infty \text{ for } x \to \partial(S^\circ)$$

Log Barrier Function

For an feasible area S defined by inequality constraints $h : \mathbb{R}^N \to \mathbb{R}^Q$:

$$S := \{x \in \mathbb{R}^N \mid h(x) \leq 0\}$$

log barrier function:

$$B(x) := -\sum_{q=1}^{Q} \log(-h_q(x))$$

convex and twice differentiable:

$$\nabla B(x) = -\sum_{q=1}^{Q} \frac{1}{h_q(x)} \nabla h_q(x)$$

$$\nabla^2 B(x) = \sum_{q=1}^{Q} \frac{1}{(h_q(x))^2} \nabla h_q(x)(\nabla h_q(x))^T - \frac{1}{h_q(x)} \nabla^2 h_q(x)$$

Inverse Barrier Function

For an feasible area S defined by inequality constraints $h : \mathbb{R}^N \rightarrow \mathbb{R}^Q$:

$$S := \{x \in \mathbb{R}^N \mid h(x) \leq 0\}$$

inverse barrier function:

$$B(x) := -\sum_{q=1}^{Q} \frac{1}{h_q(x)}$$

convex and twice differentiable:

$$\nabla B(x) = \sum_{q=1}^{Q} \frac{1}{(h_q(x))^2} \nabla h_q(x)$$

$$\nabla^2 B(x) = \sum_{q=1}^{Q} \frac{-2}{(h_q(x))^3} \nabla h_q(x) (\nabla h_q(x))^T + \frac{1}{(h_q(x))^2} \nabla^2 h_q(x)$$

Barrier Methods / Generic Algorithm

min-barrier$(f, B, x^{(0)}, c, \epsilon, K)$:
  for $k := 1, \dots, K$:
    $x^{(k)} := \min(f + c^{(k)} B, x^{(k-1)})$
    if $||x^{(k)} - x^{(k-1)}|| < \epsilon$:
      return $x^{(k)}$
  return "not converged"

where

- $f : \mathbb{R}^N \to \mathbb{R}$ objective function
- $B : \mathbb{R}^N \to \mathbb{R}$ barrier function (encoding inequality constraints)
- $x^{(0)} \in \mathbb{R}^N$ strictly feasible starting point, i.e., $B(x^{(0)}) < \infty$
- $c \in (\mathbb{R}^+)^*$: barrier weights, $c^{(k)} \to 0$
- min: unconstrained minimization method

Barrier Methods / Log Barrier Algorithm

min-barrier-log$(f, h, x^{(0)}, c, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $x^{(k)} := \min(f - c^{(k)} \sum_{q=1}^{Q} \log(-h_q), x^{(k-1)})$
    if $||x^{(k)} - x^{(k-1)}|| < \epsilon$:
      return $x^{(k)}$
  return "not converged"

where

- $f : \mathbb{R}^N \to \mathbb{R}$ objective function
- $h : \mathbb{R}^N \to \mathbb{R}^Q$ inequality constraints
- $x^{(0)} \in \mathbb{R}^N$ strictly feasible starting point, i.e., $h(x^{(0)}) < 0$
- $c \in (\mathbb{R}^+)^*$: barrier weights, $c^{(k)} \to 0$
- min: unconstrained minimization method

Remarks

- The inner minimization step is called centering step.
- It is usually accomplished using Newton's method.
- See for a better stopping criterion in section 4.

Equality Constraints

- equality constraints can be passed through to the inner problem:

$$
\begin{aligned}
x = \arg\min_{x \in \mathbb{R}^N} \; & f(x) \\
\text{s.t. } & g(x) = 0 \\
& h(x) \leq 0
\end{aligned}
\qquad \Longleftrightarrow \qquad
\begin{aligned}
x &= \lim x^{(k)}, \quad c^{(k)} \rightarrow 0 \\
x^{(k)} &:= \arg\min_{x \in S^\circ} \; \tilde{f}_{c^{(k)}}(x) \\
& \text{s.t. } g(x) = 0 \\
\tilde{f}_c(x) &:= f(x) + cB(x) \\
S^\circ &:= \{x \in \mathbb{R}^N \mid h(x) < 0\}
\end{aligned}
$$

with B a barrier function for inequality constraints h.

- the inner minimization method then has to be able to cope with equality constraints.

Outline

Idea

- search unconstrained in all of $\mathbb{R}^N$.
  - penalize infeasible points
    by adding a penalty function P to the objective
    - the penalty P is zero for feasible points, non-zero for infeasible points.

- iteratively increase the weight c of the penalty.
  - iterates $x^{(k)}$ converge to the optimum x^*,
    possibly on the border of the feasible area.

- applicable to both, equality and inequality constraints,
  but usually there are no inequality constraints.

Idea

For $f : S \to \mathbb{R}$ and $S \subseteq \mathbb{R}^N$:

$$x = \arg\min_{x \in S} \; f(\mathbf{x}) \qquad \Longleftrightarrow \qquad x = \lim x^{(k)}, \quad c^{(k)} \to \infty$$

$$x^{(k)} := \arg\min_{x \in \mathbb{R}^N} \; \tilde{f}_{c^{(k)}}(\mathbf{x})$$

$$\tilde{f}_c(x) := f(\mathbf{x}) + cP(\mathbf{x})$$

with a penalty function

$$\begin{aligned} P :& \mathbb{R}^N \to \mathbb{R} \\ (i)&P \text{ continuous} \\ (ii)&P(x) \geq 0 \\ (iii)&P(x) = 0 \Leftrightarrow x \in S \end{aligned}$$

Quadratic Penalty Function

For an feasible area S defined by equality constraints $g : \mathbb{R}^N \rightarrow \mathbb{R}^P$:

$$S := \{x \in \mathbb{R}^N \mid g(x) = 0\}$$

quadratic penalty function:

$$P(x) := \sum_{p=1}^{P} (g_p(x))^2$$

convex and twice differentiable:

$$\nabla P(x) = 2 \sum_{p=1}^{P} g_p(x) \nabla g_p(x)$$

$$\nabla^2 P(x) = 2 \sum_{p=1}^{P} \nabla g_p(x) (\nabla g_p(x))^T + g_p(x) \nabla^2 g_p(x)$$

Penalty Methods / Generic Algorithm

min-penalty$(f, P, x^{(0)}, c, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $x^{(k)} := \min(f + c^{(k)} P, x^{(k-1)})$
    if $||x^{(k)} - x^{(k-1)}|| < \epsilon$:
      return $x^{(k)}$
  return "not converged"

where

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ objective function
- $P : \mathbb{R}^N \rightarrow \mathbb{R}$ penalty function (encoding equality constraints)
- $x^{(0)} \in \mathbb{R}^N$ starting point (possibly infeasible)
- $c \in (\mathbb{R}^+)^*$: penalty weights, $c^{(k)} \rightarrow \infty$
- min: unconstrained minimization method

Penalty Methods / Quadratic Penalty Algorithm

min-penalty-quad$(f, g, x^{(0)}, c, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $x^{(k)} := \min(f + c^{(k)} \sum_{p=1}^{P} (g_p(x))^2, x^{(k-1)})$
    if $||x^{(k)} - x^{(k-1)}|| < \epsilon$:
      return $x^{(k)}$
  return "not converged"

where

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ objective function
- $g : \mathbb{R}^N \rightarrow \mathbb{R}^P$ equality constraints
- $x^{(0)} \in \mathbb{R}^N$ starting point (possibly infeasible)
- $c \in (\mathbb{R}^+)^*$: penalty weights, $c^{(k)} \rightarrow \infty$
- min: unconstrained minimization method

Inequality Constraints

- inequality constraints $h(x) \leq 0$ can be represented as (additional) equality constraints:

$$h(x) \leq 0 \quad \iff \quad h_q^+(x) := \max\{0, h_q(x)\} = 0, \quad q = 1, \ldots, Q$$

- the quadratic barrier function for h^+ is differentiable with a continuous gradient:

$$B(x) := \sum_{q=1}^{Q} (h_q^+(x))^2$$

$$\nabla B(x) = \sum_{q=1}^{Q} 2h_q^+(x) \begin{cases} \nabla h_q(x), & \text{if } h_q(x) \geq 0 \\ 0, & \text{else} \end{cases} = 2h_q^+(x) \nabla h_q(x)$$

- but the second derivative usually is not continuous on the border (where $h_q(x) = 0$).

Outline

Sequential Subproblems

Analysis for

- general inequality constraints $h(\mathbf{x}) \leq 0$
- affine equality constraints $A\mathbf{x} - \mathbf{a} = 0$

$$
\begin{aligned}
(v1) \qquad & \text{minimize } f(\mathbf{x}) \\
& \text{s.t. } h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \\
& \qquad A\mathbf{x} - \mathbf{a} = 0 \\
(v2) \qquad & \text{minimize } f(\mathbf{x}) + cB(\mathbf{x}), \quad c \to 0 \\
& \text{s.t. } A\mathbf{x} - \mathbf{a} = 0 \\
(v3) \qquad & \text{minimize } tf(\mathbf{x}) + B(\mathbf{x}), \quad t \to \infty \\
& \text{s.t. } A\mathbf{x} - \mathbf{a} = 0
\end{aligned}
$$

Central Path

Given our ICM problem

$$\begin{aligned} \text{minimize} \quad & tf(\mathbf{x}) + B(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} - \mathbf{a} = 0 \end{aligned}$$

let $\mathbf{x}^*(t)$ be its the solution for a given $t > 0$

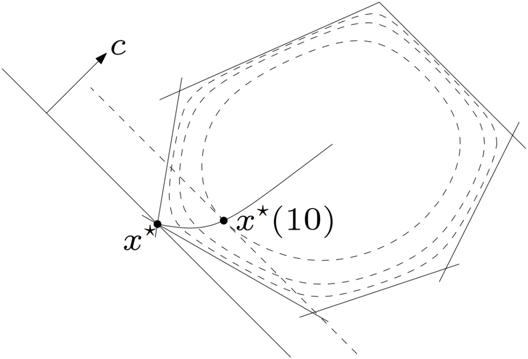
Definition
The **Central Path** associated with an ICM problem is the set of points $\mathbf{x}^*(t)$, $t > 0$, which are called **central points**

Central Path - Example

Central Path for a Linear Program

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i, \\ & i = 1, \ldots, 6 \end{array}$$

$\mathbf{c}^T\mathbf{x} = \mathbf{c}^T x^*(t)$ is tangent to the level curve of B through $\mathbf{x}^*(t)$



(From Stephen Boyd's Lecture Notes)

Central Path

Given our ICM problem

$$\begin{array}{rl} \text{minimize} & tf(\mathbf{x}) + B(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} - \mathbf{a} = 0 \end{array}$$

A point $\mathbf{x}^*(t)$ on the central path is strictly feasible, i.e., satisfies

$$A\mathbf{x}^*(t) = \mathbf{a}, \qquad h_q(\mathbf{x}^*(t)) < 0, \quad q = 1, \ldots, Q$$

and there exists a $\hat{\nu} \in \mathbb{R}^P$ such that the following holds:

$$\begin{aligned} 0 &= t\nabla f(\mathbf{x}^*(t)) + \nabla B(\mathbf{x}^*(t)) + A^T\hat{\nu} \\ &= t\nabla f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \frac{1}{-h_q(\mathbf{x}^*(t))} \nabla h_q(\mathbf{x}^*(t)) + A^T\hat{\nu} \end{aligned}$$

Dual Points from Central Path

$$0 = t\nabla f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \frac{1}{-h_q(\mathbf{x}^*(t))} \nabla h_q(\mathbf{x}^*(t)) + A^T \hat{\nu}$$

$$= \nabla f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \frac{1}{-th_q(\mathbf{x}^*(t))} \nabla h_q(\mathbf{x}^*(t)) + \frac{1}{t} A^T \hat{\nu}$$

If we define:

$$\lambda_q(t) := -\frac{1}{th_q(\mathbf{x}^*(t))},\ q = 1, \ldots, Q,\ \nu^*(t) = \frac{\hat{\nu}}{t}$$

We can rewrite:

$$\nabla f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \lambda_q(t) \nabla h_q(\mathbf{x}^*(t)) + A^T \nu^*(t) = 0$$

Minimizing the Lagrangian

From the last slide:

$$\nabla f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \lambda_q(t) \nabla h_q(\mathbf{x}^*(t)) + A^T \nu^*(t) = 0$$

we can see that this is the first order condition for the lagrangian:

$$L(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q h_q(\mathbf{x}) + \nu^T (A\mathbf{x} - \mathbf{a})$$

- $\mathbf{x}^*(t)$ minimizes the lagrangian for $\lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$
- Thus $\lambda^*(t), \nu^*(t)$ is a dual feasible pair.

The dual function

The dual function $g(\lambda^*(t), \nu^*(t))$ is finite and

$$g(\lambda^*(t), \nu^*(t)) = f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \lambda_q(t) h_q(\mathbf{x}^*(t)) + \nu^*(t)^T (A\mathbf{x}^*(t) - a)$$

$$= f(\mathbf{x}^*(t)) + \sum_{q=1}^{Q} \overbrace{-\frac{1}{t h_q(\mathbf{x}^*(t))}}^{\lambda_q(t)} h_q(\mathbf{x}^*(t)) + \nu^*(t)^T \overbrace{(A\mathbf{x}^*(t) - a)}^{A\mathbf{x}^*(t) = a}$$

$$= f(\mathbf{x}^*(t)) - \frac{Q}{t}$$

As an important consequence of this we have that:

$$f(\mathbf{x}^*(t)) - p^* \leq Q/t$$

which confirms that $\mathbf{x}^*(t)$ converges to an optimal point as $t \to \infty$

Centrality Conditions and the KKT Conditions

In order for a point $\mathbf{x}$ to be a central point, i.e. $\mathbf{x} = \mathbf{x}^*(t)$, there must exist λ, ν such that:

$$A\mathbf{x} = \mathbf{a}, \quad h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q$$

$$\lambda \geq 0$$

$$\nabla f(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \nabla h_q(\mathbf{x}) + A^T \nu = 0$$

$$-\lambda_q h_q(\mathbf{x}) = \frac{1}{t}, \quad q = 1, \ldots, Q$$

- Thus, $\mathbf{x}^*(t)$ almost fulfills the KKT conditions.
  - complementary condition $-\lambda_q h_q(\mathbf{x}) = 0$ only holds approximately $(= 1/t)$

Stopping Criterion

- as stopping criterion, simply

$$\frac{Q}{t} \leq \epsilon, \quad t \to \infty$$

  or equivalently

$$Qc \leq \epsilon, \quad c \to 0$$

  can be used.
- Why solving sequential problems? Why not just solve a single problem with a sufficiently small c? E.g.,

$$c := \frac{\epsilon}{Q}$$

  - It does not work well for large scale problems.
  - It does not work well for small accuracy ϵ.
  - It needs a "good" starting point.
- Trade-off about the schedule of c:
  - the smaller c, the fewer centering steps, but the more Newton steps / centering step
  - can be adaptively controlled.

Outline

Convergence Analysis

Assume that $tf + B$ can be minimized by Newton's method for $t = t^{(0)}, \mu t^{(0)}, \mu^2 t^{(0)}, \ldots$, the t in the k-th outer step is

$$t^{(k)} = \mu^k t^{(0)}$$

From this, it follows that, in the k-th outer step, the duality gap is

$$\frac{Q}{\mu^k t^{(0)}}$$

Convergence Analysis

Then the number of outer iterations k^* needed to achieve accuracy ϵ is

$$\begin{aligned}
\epsilon &= \frac{Q}{\mu^{k^*} t^{(0)}} \\
\mu^{k^*} &= \frac{Q}{\epsilon t^{(0)}} \\
\log(\mu^{k^*}) &= \log(\frac{Q}{\epsilon t^{(0)}}) \\
k^* \log(\mu) &= \log(\frac{Q}{\epsilon t^{(0)}}) \\
k^* &= \frac{\log(\frac{Q}{\epsilon t^{(0)}})}{\log(\mu)}
\end{aligned}$$

Convergence Analysis

The **number of outer iterations** is exactly:

$$\left\lceil \frac{\log\left(\frac{Q}{\epsilon t^{(0)}}\right)}{\log \mu} \right\rceil$$

plus the initial step to compute $\mathbf{x}^*(t^{(0)})$

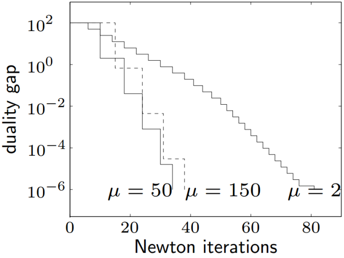
The **inner problem**

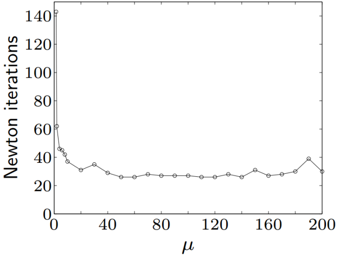
$$\text{minimize} \quad tf(\mathbf{x}) + B(\mathbf{x})$$

is solved by Newton's method (see convergence analysis for it)

Examples

Inequality form Linear Program ($m = 100$ inequalities, $n = 50$ variables)





(From Stephen Boyd's Lecture Notes)

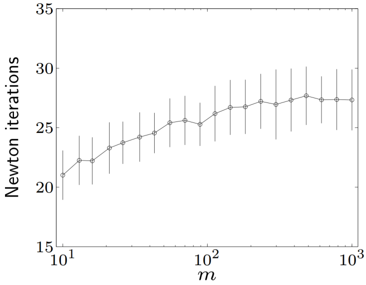
- starts with $\mathbf{x}$ on central path ($t^{(0)} = 1$, duality gap 100)
- terminates when $t = 10^8$ (gap 10^{-6})
- centering uses Newton's method with backtracking
- total number of Newton iterations not very sensitive for $\mu \geq 10$

Examples

Family of Linear Programs ($A \in \mathbb{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & A^T x \leq b, \quad x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m solve 100 randomly generated instances



Outline

Feasibility and Phase I Method

- The barrier method requires a strictly feasible starting point $\mathbf{x}^{(0)}$.
- Phase I denotes the computation of such a point (or the constraints are found to be infeasible).
- The barrier method algorithm then starts from $\mathbf{x}^{(0)}$ (called phase II stage).

Basic Phase I Method

Find strictly feasible $\mathbf{x}$ for constraints

$$h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q, \quad A\mathbf{x} - \mathbf{a} = 0 \tag{1}$$

Phase I method for target variables $\mathbf{x} \in \mathbb{R}^N$ and $s \in \mathbb{R}$:

$$\begin{aligned} \text{minimize} \quad & s \\ \text{subject to} \quad & h_q(\mathbf{x}) \leq s, \quad q = 1, \ldots, Q \\ & A\mathbf{x} - \mathbf{a} = 0 \end{aligned} \tag{2}$$

- for (2), a strictly feasible starting point is easy to compute:
  - compute $x^{(0)}$ with $Ax^{(0)} - a = 0$
  - $s^{(0)} := \max_{q=1,\ldots,Q} h_q(x^{(0)}) + \epsilon, \quad \epsilon > 0$
- if $\mathbf{x}, s$ is feasible, with $s < 0$, then $\mathbf{x}$ is strictly feasible for (1)
- if the optimal value s^* of (2) is positive, then problem (1) is infeasible
- if $s^* = 0$ and attained, then problem (1) is feasible (but not strictly)
- if $s^* = 0$ and not attained, then problem (1) is infeasible

Sum of Infeasibilities Phase I Method

For target variables $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{s} \in \mathbb{R}^Q$:

$$\begin{aligned} \text{minimize} \quad & \mathbf{1}^T \mathbf{s} \\ \text{subject to} \quad & \mathbf{s} \geq 0 \\ & h_q(\mathbf{x}) \leq s_q, \quad q = 1, \ldots, Q \\ & A\mathbf{x} - \mathbf{a} = 0 \end{aligned}$$

Further Readings

- Barrier methods:
  - [Boyd and Vandenberghe, 2004, ch. 11]
  - [Griva et al., 2009, ch. 16]
  - [Luenberger and Ye, 2008, ch. 13]
  - [Nocedal and Wright, 2006, ch. 19.6]

- Penalty methods:
  - [Griva et al., 2009, ch. 16]
  - [Luenberger and Ye, 2008, ch. 13]
  - [Nocedal and Wright, 2006, ch. 17.1–2]

References I

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ Press, 2004.

Igor Griva, Stephen G. Nash, and Ariela Sofer. *Linear and nonlinear optimization*. Society for Industrial and Applied Mathematics, 2009.

David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*. Springer, 2008. Fourth edition 2015.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2006.