Part 6. Newton's Method

Math 126 Winter 18

Date of current version: February 28, 2018

Abstract This note studies Newton's methods. Many parts of this note are based on the chapters [1, Chapter 5] [2, Chapters 9-11] [3, Chapters 1,4].

Please email me if you find any typos or errors.

1 Newton's Method (see [1, Chapter 5] [2, Chapter 9] [3, Chapter 1])

Consider the following unconstrained minimization problem:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}), \tag{1.1}$$

where f is twice continuously differentiable.

1.1 Newton's Method

The update of the Newton's method is given by

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{x} - \boldsymbol{x}_k \rangle + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^\top \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) \right\}, \tag{1.2}$$

which reduces to the following when we further assume that $\nabla^2 f(\boldsymbol{x}_k)$ is positive definite:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - [\nabla^2 f(\boldsymbol{x}_k)]^{-1} \nabla f(\boldsymbol{x}_k). \tag{1.3}$$

Remark 1.1 Newton's direction $\boldsymbol{d} = -[\nabla^2 f(\boldsymbol{x})]^{-1} \nabla f(\boldsymbol{x}) \neq \boldsymbol{0}$ is not necessarily a descent direction, *i.e.*,

$$\nabla f(\boldsymbol{x})^\top \boldsymbol{d} = -\nabla f(\boldsymbol{x})^\top \nabla^2 f(\boldsymbol{x}) \nabla f(\boldsymbol{x}) \not< 0. \tag{1.4}$$

If $\nabla^2 f(\boldsymbol{x})$ is positive definite, then Newton's direction is a descent direction.

At the very least, Newton's method requires that $\nabla^2 f(\boldsymbol{x})$ is positive definite for every $\boldsymbol{x} \in \mathbb{R}^d$, which in particular implies that there exists a unique optimal solution $\boldsymbol{x}_*$. However, this is not enough to guarantee convergence, as the following example illustrates.

Example 1.1 Consider $f(x) = \sqrt{1 + x^2}$ defined over the real line. The minimizer of f over $\mathbb{R}$ is $x_* = 0$. The first and second derivatives of f are

$$f'(x) = \frac{x}{\sqrt{1 + x^2}}, \quad f''(x) = \frac{1}{(1 + x^2)^{\frac{3}{2}}}.$$

Donghwan Kim
Dartmouth College
E-mail: donghwan.kim@dartmouth.edu

The update of Newton's method hast the form

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = x_k - x_k(1 + x_k^2) = -x_k^3.$$

We therefore see that for $|x_0| \geq 1$ the method diverges and that for $|x_0| < 1$ the method converges very rapidly to the solution $x_* = 0$.

Theorem 1.1 *(quadratic local convergence of Newton's method) Let f be a twice continuously differentiable function defined over $\mathbb{R}^d$. Assume that*

- *there exists $\mu > 0$ for which $\nabla^2 f(\boldsymbol{x}) \succeq \mu \boldsymbol{I}$ for any $\boldsymbol{x} \in \mathbb{R}^d$,*
- *there exists $M > 0$ for which $||\nabla^2 f(\boldsymbol{x}) - \nabla^2 f(\boldsymbol{y})||_2 \leq M||\boldsymbol{x} - \boldsymbol{y}||_2$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.*

Let $\{\boldsymbol{x}_k\}_{k\geq 0}$ be the sequence generated by Newton's method, and let $\boldsymbol{x}_*$ be unique minimizer of f over $\mathbb{R}^d$. Then

$$||\boldsymbol{x}_{k+1} - \boldsymbol{x}_*||_2 \leq \frac{M}{2\mu}||\boldsymbol{x}_k - \boldsymbol{x}_*||_2^2. \tag{1.5}$$

In addition, if $||\boldsymbol{x}_0 - \boldsymbol{x}_*||_2 \leq \frac{\mu}{M}$, then

$$||\boldsymbol{x}_k - \boldsymbol{x}_*||_2 \leq \frac{2\mu}{M}\left(\frac{1}{2}\right)^{2^k}. \tag{1.6}$$

Proof *By the fundamental theorem of calculus, we have*

$$\begin{aligned}
\boldsymbol{x}_{k+1} - \boldsymbol{x}_* &= \boldsymbol{x}_k - [\nabla^2 f(\boldsymbol{x}_k)]^{-1}\nabla f(\boldsymbol{x}_k) - \boldsymbol{x}_* \\
&= \boldsymbol{x}_k - \boldsymbol{x}_* + [\nabla^2 f(\boldsymbol{x}_k)]^{-1}(\nabla f(\boldsymbol{x}_*) - \nabla f(\boldsymbol{x}_k)) \\
&= \boldsymbol{x}_k - \boldsymbol{x}_* + [\nabla^2 f(\boldsymbol{x}_k)]^{-1}\int_0^1 [\nabla^2 f(\boldsymbol{x}_k + t(\boldsymbol{x}_* - \boldsymbol{x}_k))](\boldsymbol{x}_* - \boldsymbol{x}_k)dt \\
&= [\nabla^2 f(\boldsymbol{x}_k)]^{-1}\int_0^1 [\nabla^2 f(\boldsymbol{x}_k + t(\boldsymbol{x}_* - \boldsymbol{x}_k)) - \nabla^2 f(\boldsymbol{x}_k)](\boldsymbol{x}_* - \boldsymbol{x}_k)dt.
\end{aligned}$$

Then,

$$\begin{aligned}
||\boldsymbol{x}_{k+1} - \boldsymbol{x}_*||_2 &\leq ||[\nabla^2 f(\boldsymbol{x}_k)]^{-1}||_2 \left\|\int_0^1 [\nabla^2 f(\boldsymbol{x}_k + t(\boldsymbol{x}_* - \boldsymbol{x}_k)) - \nabla^2 f(\boldsymbol{x}_k)](\boldsymbol{x}_* - \boldsymbol{x}_k)dt\right\|_2 \\
&\leq ||[\nabla^2 f(\boldsymbol{x}_k)]^{-1}||_2 \int_0^1 \left\|[\nabla^2 f(\boldsymbol{x}_k + t(\boldsymbol{x}_* - \boldsymbol{x}_k)) - \nabla^2 f(\boldsymbol{x}_k)](\boldsymbol{x}_* - \boldsymbol{x}_k)\right\|_2 dt \\
&\leq ||[\nabla^2 f(\boldsymbol{x}_k)]^{-1}||_2 \int_0^1 \left\|\nabla^2 f(\boldsymbol{x}_k + t(\boldsymbol{x}_* - \boldsymbol{x}_k)) - \nabla^2 f(\boldsymbol{x}_k)\right\|_2 \|\boldsymbol{x}_* - \boldsymbol{x}_k\|_2 dt \\
&\leq \frac{1}{\mu}\int_0^1 Mt||\boldsymbol{x}_k - \boldsymbol{x}_*||_2^2 dt \\
&= \frac{M}{2\mu}||\boldsymbol{x}_k - \boldsymbol{x}_*||_2^2
\end{aligned}$$

We next prove the second inequality (1.6) by induction. Note that for $k = 0$, we assumed that

$$||\boldsymbol{x}_0 - \boldsymbol{x}_*||_2 \leq \frac{\mu}{M} = \frac{2\mu}{M}\left(\frac{1}{2}\right)^{2^0}.$$

Assume that (1.6) holds for an integer k, that is $||\boldsymbol{x}_k - \boldsymbol{x}_*||_2 \leq \frac{2\mu}{M}\left(\frac{1}{2}\right)^{2^k}$; we will show it holds for $k+1$. By (1.5) we have

$$||\boldsymbol{x}_{k+1} - \boldsymbol{x}_*||_2 \leq \frac{M}{2\mu}||\boldsymbol{x}_k - \boldsymbol{x}_*||_2^2 \leq \frac{M}{2\mu}\left(\frac{2\mu}{M}\left(\frac{1}{2}\right)^{2^k}\right)^2 = \frac{2\mu}{M}\left(\frac{1}{2}\right)^{2^{k+1}}.$$

□

Remark 1.2 An iterative method is called locally convergent if the generated sequence converges to an optimal point $\boldsymbol{x}_*$ given that the initial point $\boldsymbol{x}_0$ is close enough to $\boldsymbol{x}_*$.

1.2 Damped Newton's Method

Newton's method does not guarantee descent of the function values even when the Hessian is positive definite, similar to a gradient method with step size $s_k = 1$, *i.e.* $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \nabla f(\boldsymbol{x}_k)$. This can be fixed by introducing a step size chosen by a certain line search, leading to the following damped Newton's method.

Algorithm 1 Damped Newton's Method

```
Input: x_0 ∈ R^d.
for k ≥ 0 do
    Compute the Newton direction d_k, which is the solution to the linear system ∇^2 f(x_k) d_k = −∇f(x_k).
    Choose a step size s_k > 0 using a backtracking line search.
    x_k = x_k + s_k d_k.
    If a stopping criteria is satisfied, then stop.
```

Remark 1.3 Backtracking line search: starting from an initial $s > 0$, repeat $s \leftarrow \beta s$ until the following sufficient decrease condition is satisfied:

$$f(\boldsymbol{x} + s\boldsymbol{d}) < f(\boldsymbol{x}) + \alpha s \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$$

with parameters $\alpha \in (0, 1)$ and $\beta \in (0, 1)$.

There exists constants $\eta \in \left(0, \frac{\mu^2}{M}\right)$, and $\gamma > 0$ such that the following holds; specifically, $0 < \eta \leq \min\left\{\frac{\mu^2}{M}, 3(1-2\alpha)\frac{\mu^2}{M}\right\}$ and $\gamma = \alpha\beta\eta^2\frac{\mu}{L^2}$ for a constant L satisfying $\nabla^2 f(\boldsymbol{x}) \preceq L\boldsymbol{I}$ for any $\boldsymbol{x} \in \mathbb{R}^d$.

- (damped Newton phase) If $||\nabla f(\boldsymbol{x}_k)||_2 \geq \eta$, then

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \leq -\gamma. \tag{1.7}$$

- (quadratically convergent phase) If $||\nabla f(\boldsymbol{x}_k)||_2 < \eta$, then the backtracking line search selects $s_k = 1$ and

$$||\nabla f(\boldsymbol{x}_{k+1})||_2 \leq \frac{M}{2\mu}^2 ||\nabla f(\boldsymbol{x}_k)||_2^2 \tag{1.8}$$

Remark 1.4 Unlike the gradient method that is affected by changes of coordinates, Newton's method is independent of affine changes of coordinates. Suppose $\boldsymbol{T} \in \mathbb{R}^{d\times d}$ is nonsingular, and define $\bar{f}(\boldsymbol{y}) = f(\boldsymbol{T}\boldsymbol{y})$. If we use Newton's method (with the same backtracking parameters) to minimize $\bar{f}$, starting from $\boldsymbol{y}_0 = \boldsymbol{T}^{-1}\boldsymbol{x}_0$, then we have $\boldsymbol{T}\boldsymbol{y}_k = \boldsymbol{x}_k$ for all k.

1.3 Self-concordance

Previous convergence analysis involves the constants μ, M, and L, which are difficult to estimate in practice. In addition, even though Newton's method is affinely invariant, the previous convergence analysis is not affinely invariant.

Consider the following assumption, called self-concordance, that does not depend on any unknown constants (such as μ, M, and L) and affine changes of coordinates.

Definition 1.1 A convex function $f : \mathbb{R} \to \mathbb{R}$ is *self-concordant* if

$$|f'''(x)| \leq \kappa f''(x)^{\frac{3}{2}} \tag{1.9}$$

for all $x \in \operatorname{dom} f$, where κ is some positive constant. $\kappa = 2$ is a standard choice.

A function $f : \mathbb{R}^d \to \mathbb{R}$ is self-concordant if it is self-concordant along every line in its domain, *i.e.*, if the function $\tilde{f}(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ is a self-concordant function of t for all $\boldsymbol{x} \in \operatorname{dom} f$ and for all $\boldsymbol{v}$.

Remark 1.5 The self-concordant functions include many of the logarithmic barrier functions that play an important role in interior point methods for solving convex optimization problems.

Example 1.2 Self-concordant functions.

- Negative logarithm $f(x) = -\log x$: Using $f''(x) = \frac{1}{x^2}$, $f'''(x) = -\frac{2}{x^3}$, we have

$$\frac{|f'''(x)|}{2f''(x)^{\frac{3}{2}}} = 1.$$

- Negative entropy plus negative logarithm $f(x) = x\log x - \log x$

Self-concordance is preserved under

- scaling by a factor $a \geq 1$,
- addition,
- composition with affine function,
- composition with logarithm: Let $g : \mathbb{R} \to \mathbb{R}$ be a convex function with $\operatorname{dom} g = \mathbb{R}_{++}$ and

$$|g'''(x)| \leq 3\frac{g''(x)}{x} \tag{1.10}$$

for all x. Then

$$f(x) = -\log(-g(x)) - \log x \tag{1.11}$$

is self-concordant on $\{x : x > 0,\ g(x) < 0\}$

Example 1.3 Self-concordant functions.

- Log barrier for linear inequalities $f(\boldsymbol{x}) = -\sum_{i=1}^{m} \log\big(b_i - \boldsymbol{a}_i^\top \boldsymbol{x}\big)$ on $\{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} < b_i,\ i = 1, \ldots, m\}$
- Log-determinant $f(\boldsymbol{X}) = -\log \mathsf{det}\, \boldsymbol{X}$ on $\mathbb{S}_{++}^d$
- $f(\boldsymbol{x}, y) = -\log\big(y^2 - \boldsymbol{x}^\top \boldsymbol{x}\big)$ on $\{(\boldsymbol{x}, y) : ||\boldsymbol{x}||_2 \leq y\}$.

For strictly convex self-concordant function, we obtain bounds in terms of the Newton decrement

$$\lambda(\boldsymbol{x}) = \sqrt{\nabla f(\boldsymbol{x})^\top [\nabla^2 f(\boldsymbol{x})]^{-1} \nabla f(\boldsymbol{x})}. \tag{1.12}$$

There exists constants $\eta \in (0, 1/4]$ and $\gamma > 0$ (that depend only on the backtracking line search parameters α and β) such that the following holds.

- (damped Newton phase) If $\lambda(\boldsymbol{x}_k) > \eta$, then

$$f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \leq -\gamma. \tag{1.13}$$

- (quadratically convergent phase) If $\lambda(\boldsymbol{x}_k) \leq \eta$, then the backtracking line search selects $s_k = 1$ and

$$\lambda(\boldsymbol{x}_{k+1}) \leq 2\lambda(\boldsymbol{x}_k)^2 \tag{1.14}$$

2 Newton's Method with Equality Constraints (see [2, Chapter 10])

Consider the equality constrained convex minimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \; & f(\boldsymbol{x}) \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{2.1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is convex and twice continuously differentiable, and $\boldsymbol{A} \in \mathbb{R}^{p\times d}$ with $\operatorname{rank}(\boldsymbol{A}) = p < d$.

To derive the Newton direction $\Delta\boldsymbol{x}_{\text{nt}}$ at the feasible point $\hat{\boldsymbol{x}}$, we solve the following second-order Taylor approximation near $\hat{\boldsymbol{x}}$:

$$\begin{aligned} \min_{\boldsymbol{v}\in\mathbb{R}^d} \; & \left\{ \hat{f}(\hat{\boldsymbol{x}}+\boldsymbol{v}) \equiv f(\hat{\boldsymbol{x}}) + \langle \nabla f(\hat{\boldsymbol{x}}),\, \boldsymbol{v}\rangle + \frac{1}{2}\boldsymbol{v}^\top \nabla^2 f(\hat{\boldsymbol{x}})\boldsymbol{v} \right\} \\ \text{subject to } & \boldsymbol{A}(\hat{\boldsymbol{x}}+\boldsymbol{v}) = \boldsymbol{b}. \end{aligned} \tag{2.2}$$

Recall that a point $\boldsymbol{v}_* \in \mathbb{R}^d$ is optimal for (2.2) if and only there is a $\boldsymbol{\mu}_* \in \mathbb{R}^p$ such that

$$\nabla_{\boldsymbol{v}} \hat{f}(\hat{\boldsymbol{x}}+\boldsymbol{v}_*) + \boldsymbol{A}^\top \boldsymbol{\mu}_* = \nabla f(\hat{\boldsymbol{x}}) + \nabla^2 f(\hat{\boldsymbol{x}})\boldsymbol{v}_* + \boldsymbol{A}^\top \boldsymbol{\mu}_* = \boldsymbol{0}, \quad \boldsymbol{A}(\hat{\boldsymbol{x}}+\boldsymbol{v}_*) = \boldsymbol{b}, \tag{2.3}$$

which is a linearized approximation of the optimality conditions of (2.1) near $\hat{\boldsymbol{x}}$:

$$\nabla f(\boldsymbol{x}_*) + \boldsymbol{A}^\top \boldsymbol{\mu}_* = \boldsymbol{0}, \quad \boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{b}. \tag{2.4}$$

Using $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, the Newton direction $\Delta\boldsymbol{x}_{\text{nt}} = \boldsymbol{v}_*$ can be simply characterized by

$$\underbrace{\begin{pmatrix} \nabla^2 f(\hat{\boldsymbol{x}}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix}}_{\text{KKT matrix}} \begin{pmatrix} \boldsymbol{v}_* \\ \boldsymbol{\mu}_* \end{pmatrix} = \begin{pmatrix} -\nabla f(\hat{\boldsymbol{x}}) \\ \boldsymbol{0} \end{pmatrix}. \tag{2.5}$$

The Newton direction is defined only at points for which the KKT matrix is nonsingular.

On the other hand, one can parameterize the feasible set $\{\boldsymbol{x} \,:\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ as

$$\{\boldsymbol{x} \,:\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} = \{\boldsymbol{F}\boldsymbol{z} + \tilde{\boldsymbol{x}} \,:\, \boldsymbol{z} \in \mathbb{R}^{d-p}\} \tag{2.6}$$

with a feasible point $\tilde{\boldsymbol{x}}$ that satisfies $\boldsymbol{A}\tilde{\boldsymbol{x}} = \boldsymbol{b}$ and a matrix $\boldsymbol{F} \in \mathbb{R}^{d\times(d-p)}$ whose range is the nullspace of $\boldsymbol{A}$. Then, one can eliminate the equality constraints of the problem (2.1) as

$$\min_{\boldsymbol{z}\in\mathbb{R}^{d-p}} \left\{ \tilde{f}(\boldsymbol{z}) \equiv f(\boldsymbol{F}\boldsymbol{z} + \tilde{\boldsymbol{x}}) \right\}. \tag{2.7}$$

The corresponding Newton's direction is

$$\Delta\boldsymbol{z}_{\text{nt}} = -[\nabla^2 \tilde{f}(\boldsymbol{z})]^{-1} \nabla \tilde{f}(\boldsymbol{z}) = -[\boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{F}]^{-1} \boldsymbol{F}^\top \nabla f(\boldsymbol{x}), \tag{2.8}$$

where $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{z} + \tilde{\boldsymbol{x}}$. One can show that (proof omitted)

$$\Delta\boldsymbol{x}_{\text{nt}} = \boldsymbol{F}\Delta\boldsymbol{z}_{\text{nt}}. \tag{2.9}$$

Then, starting from $\boldsymbol{x}_0 = \boldsymbol{F}\boldsymbol{z}_0 + \tilde{\boldsymbol{x}}$, the iterates $\{\boldsymbol{x}_k\}_{k\geq 0}$ and $\{\boldsymbol{z}_k\}_{k\geq 0}$ of Newton's method for (2.1) and (2.7) respectively satisfy

$$\boldsymbol{x}_k = \boldsymbol{F}\boldsymbol{z}_k + \tilde{\boldsymbol{x}}, \tag{2.10}$$

and thus the convergence analyis of Newton's method for unconstrained problem directly applies to Newton's method solving the equality constrainted problems.

What if we don't have a feasible point $\tilde{\boldsymbol{x}}$? A Newton's direction $\Delta\boldsymbol{x}$ can be found by solving

$$\begin{pmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \boldsymbol{\mu} \end{pmatrix} = -\begin{pmatrix} \nabla f(\boldsymbol{x}) \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \end{pmatrix}. \tag{2.11}$$

3 Newton's Method with Inequality Constraints (see [2, Chapter 11] [3, Chapter 4])

Consider the equality and inequality constrained convex minimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \; & f(\boldsymbol{x}) \\ \text{subject to } & g_i(\boldsymbol{x}) \le 0, \quad i = 1, \dots, m, \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{3.1}$$

where $f, g_i : \mathbb{R}^d \to \mathbb{R}$ are convex and twice continuously differentiable, and $\boldsymbol{A} \in \mathbb{R}^{p\times d}$ with $\operatorname{rank}(\boldsymbol{A}) = p < d$. We also assume that there exists $\hat{\boldsymbol{x}} \in \operatorname{dom} f$ that satisfies $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{b}$ and $g_i(\boldsymbol{x}) < 0$ for $i = 1, \dots, m$.

We approximate the problem into an equality constrained problem to which Newton's method can be applied. We first use an indication function to make the inequality constraints implict as:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \; & f(\boldsymbol{x}) + \sum_{i=1}^m I_{\mathbb{R}_-}(g_i(\boldsymbol{x})) \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned} \tag{3.2}$$

The objective function is yet non-differentiable, so Newton's method cannot be applied.

The basic idea of the barrier method, a particular interior-point method, is to approximate the indicator function $I_{\mathbb{R}_-}$ by

$$I_{\mathbb{R}_-}(u) = I_{\mathbb{R}_+}(-u) \approx \hat{I}_-(u) = -\frac{1}{t}\log(-u) \tag{3.3}$$

with $t > 0$, where $\operatorname{dom} \hat{I}_- = -\mathbb{R}_{++}$. The approximation improves as $t \to \infty$. Substituting $\hat{I}_-$ for $I_{\mathbb{R}_-}$ leads to

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \; & \left\{ f(\boldsymbol{x}) - \sum_{i=1}^m \frac{1}{t}\log(-g_i(\boldsymbol{x})) \right\} \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{3.4}$$

which can be solved by Newton's method.

The function

$$\psi(\boldsymbol{x}) = -\sum_{i=1}^m \log(-g_i(\boldsymbol{x})) \tag{3.5}$$

is called the logarithmic barrier (or interior penalty), which is convex and twice continously differentiable.

For $t > 0$, define $\boldsymbol{x}_*(t)$ as the solution of

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^d} \; & \{ t f(\boldsymbol{x}) + \psi(\boldsymbol{x}) \} \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{3.6}$$

which can be solved via Newton's method. The *central path* associated with the problem (3.6) is defined as $\{\boldsymbol{x}_*(t) \,:\, t > 0\}$, a set of central points $\boldsymbol{x}_*(t)$.

Central points $\boldsymbol{x}_*(t)$ for $t > 0$ are characterized by the following necessary and sufficient conditions: $\boldsymbol{x}_*(t)$ is strictly feasible for (3.1), *i.e.*, satisfies

$$g_i(\boldsymbol{x}_*(t)) < 0, \; i = 1, \dots, m, \quad \boldsymbol{A}\boldsymbol{x}_*(t) = \boldsymbol{b}, \tag{3.7}$$

and there exists $\hat{\boldsymbol{\mu}}$ such that

$$\begin{aligned} \boldsymbol{0} &= t\nabla f(\boldsymbol{x}_*(t)) + \nabla\psi(\boldsymbol{x}_*(t)) + \boldsymbol{A}^\top \hat{\boldsymbol{\mu}} \\ &= t\nabla f(\boldsymbol{x}_*(t)) + \sum_{i=1}^m \left(-\frac{1}{g_i(\boldsymbol{x}_*(t))} \right) \nabla g_i(\boldsymbol{x}_*(t)) + \boldsymbol{A}^\top \hat{\boldsymbol{\mu}}. \end{aligned} \tag{3.8}$$

We can then state that every central point $\boldsymbol{x}_*(t)$ yields a dual feasible point $(\boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t))$:

$$\lambda_{*,i}(t) = -\frac{1}{tg_i(\boldsymbol{x}_*(t))},\ i = 1, \ldots, m, \quad \boldsymbol{\mu}_*(t) = \frac{\hat{\boldsymbol{\mu}}}{t}, \tag{3.9}$$

and hence a lower bound on the primal optimal value p_*. Specifically, from (3.8), a central point $\boldsymbol{x}_*(t)$ minimizes the Lagragian of (3.1) for $(\boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t))$:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t)) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_{*,i}(t) g_i(\boldsymbol{x}) + \boldsymbol{\mu}_*(t)^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}). \tag{3.10}$$

Then the dual function is

$$q(\boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t)) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t)) = f(\boldsymbol{x}_*(t)) - \frac{m}{t}, \tag{3.11}$$

which implies

$$f(\boldsymbol{x}_*(t)) - p_* \leq f(\boldsymbol{x}_*(t)) - g(\boldsymbol{\lambda}_*(t), \boldsymbol{\mu}_*(t)) = \frac{m}{t}. \tag{3.12}$$

This confirms our intuition that $\boldsymbol{x}_*(t)$ converges to an optimal point as $t \to \infty$.

Algorithm 2 Barrier Method (Path-following Method)

Input: strictly feasible $\boldsymbol{x}_0 \in \mathbb{R}^d$, $t_0 > 0$, $\rho > 1$.
for $k \geq 0$ **do**
    Compute $\boldsymbol{x}_*(t_k)$ by minimizing $tf(\boldsymbol{x}) + \psi(\boldsymbol{x})$ subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ starting at $\boldsymbol{x}_k$ (via Newton's method).
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_*(t_k)$
    If a stopping criteria is satisfied, then stop.
    $t_{k+1} = \rho t_k$.

References

1. Beck, A.: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Soc. Indust. Appl. Math., Philadelphia (2014). URL http://epubs.siam.org/doi/abs/10.1137/1.9781611973655
2. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge, UK (2004). URL http://www.stanford.edu/~boyd/cvxbook.html
3. Nesterov, Y.: Introductory lectures on convex optimization: A basic course. Kluwer (2004). DOI 10.1007/978-1-4419-8853-9