164 Final Revision

Jiantong Liu

June 8, 2022

Contents

1 Basics of Set-constrained and Unconstained Optimization

Definition 4.3. A set $\Theta \subseteq \mathbb{R}^n$ is convex if for all $\mathbf{u}, \mathbf{v} \in \Omega$, the line segment between $\mathbf{u}$ and $\mathbf{v}$ is in Θ, i.e. $\alpha \mathbf{u} + (1-\alpha)\mathbf{v} \in \Omega$ for all $\mathbf{u}, \mathbf{v} \in \Theta$ and $\alpha \in (0,1)$.

Definition 6.1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subseteq \mathbb{R}^n$. A point $\mathbf{x}^* \in \Omega$ is a local minimizer of f over Ω if there exists $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \backslash \{\mathbf{x}^*\}$ and $||\mathbf{x} - \mathbf{x}^*|| < \varepsilon$. A point $\mathbf{x}^* \in \Omega$ is a global minimizer of f over Ω if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \backslash \{\mathbf{x}^*\}$.

Definition 6.2. A vector $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$, is a feasible direction at $\mathbf{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\mathbf{x} + \alpha \mathbf{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.

FONC . Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\mathbf{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we have

$$\mathbf{d}^T \nabla f(\mathbf{x}^*) \geq 0.$$

FONC Interior Case . Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\mathbf{x}^*$ is a local minimizer of f over Ω and if $\mathbf{x}^*$ is an interior point of Ω, then

$$\nabla f(\mathbf{x}^*) = \mathbf{0}.$$

SONC . Let $\Omega \subseteq \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\mathbf{x}^*$ a local minimizer of f over Ω, and $\mathbf{d}$ a feasible direction at $\mathbf{x}^*$. If $\mathbf{d}^T \nabla f(\mathbf{x}^*) = 0$, then

$$\mathbf{d}^T \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0,$$

where $\mathbf{F}$ is the Hessian of f.

SONC Interior Case . Let $\mathbf{x}^*$ be an interior point of $\Omega \subseteq \mathbb{R}^n$. If $\mathbf{x}^*$ is a local minimizer of $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, then

$$\nabla f(\mathbf{x}^*) = \mathbf{0},$$

and $\mathbf{F}(\mathbf{x}^*)$ is positive semidefinite ($\mathbf{F}(\mathbf{x}^*) \geq 0$); that is, for all $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^T \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0.$$

SOSC . Let $f \in \mathcal{C}^2$ be defined on a region in which $\mathbf{x}^*$ is an interior point. Suppose that

1. $\nabla f(\mathbf{x}^*) = \mathbf{0}$.
2. $\mathbf{F}(\mathbf{x}^*) > 0$.

Then $\mathbf{x}^*$ is a strict local minimizer.

2 Gradient's Method

2.1 Gradient Descent Algorithm

Recursive formula

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

where α_k is a positive scalar called the step size.

2.2 The Method of Steepest Descent

In particular, α_k is chosen to minimize $\varphi_k(\alpha) = f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)}))$. In other words,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)})).$$

Proposition 8.1. If $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ is a steepest descent sequence for a given function $f : \mathbb{R}^n \to \mathbb{R}$, then for each k the vector $\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ is orthogonal to the vector $\mathbf{x}^{(k+2)} - \mathbf{x}^{(k+1)}$.

Proposition 8.2. If $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ is the steepest descent sequence for $f : \mathbb{R}^n \to \mathbb{R}$ and if $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$, then $f(\mathbf{x}^{(k+)}) < f(\mathbf{x}^{(k)})$.

2.3 Method of Steepest Descent on Quadratic Function

For quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$, where $\mathbf{Q} \in \mathbb{R}^{(n \times n)}$ is a symmetric positive definite matrix, the algorithm takes the form

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \frac{(\mathbf{g}^{(k)})^T\mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^T\mathbf{Q}\mathbf{g}^{(k)}}\mathbf{g}^{(k)}$$

where $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)}) = \mathbf{Q}\mathbf{x}^{(k)} - \mathbf{b}$.

Lemma 8.1. The iterative algorithm $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k\mathbf{g}^{(k)}$ with $\mathbf{g}^{(k)} = \mathbf{Q}\mathbf{x}^{(k)} - \mathbf{b}$ satisfies

$$V(\mathbf{x}^{(k+1)}) = (1 - \gamma_k)V(\mathbf{x}^{(k)}),$$

where if $\mathbf{g}^{(k)} = \mathbf{0}$, then $\gamma_k = 1$, and if $\mathbf{g}^{(k)} \neq \mathbf{0}$, then

$$\gamma_k = \alpha_k \frac{(\mathbf{g}^{(k)})^T\mathbf{Q}\mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^T\mathbf{Q}^{-1}\mathbf{g}^{(k)}}\left(2\frac{(\mathbf{g}^{(k)})^T\mathbf{g}^{(k)}}{(\mathbf{g}^{(k)})^T\mathbf{Q}\mathbf{g}^{(k)}} - \alpha_k\right).$$

Theorem 8.1. Let $\{\mathbf{x}^{(k)}\}$ be the result of gradient algorithm $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$, then the sequence converges to $\mathbf{x}^*$ for any initial condition $\mathbf{x}^{(0)}$ if and only if $\sum\limits_{k=0}^{\infty} \gamma_k = \infty$.

Theorem 8.2. In the steepest descent algorithm, we have $\mathbf{x}^{(k)} \to \mathbf{x}^*$ for any $\mathbf{x}^{(0)}$.

Theorem 8.3. For the fixed-step-size gradient algorithm, $\mathbf{x}^{(k)} \to \mathbf{x}^*$ for $\mathbf{x}^{(0)}$ if and only if

$$0 < \alpha < \frac{2}{\lambda_{\max(\mathbf{Q})}}.$$

Theorem 8.4. In the method of steepest descent applied to the quadratic function, at every step k we have

$$V(\mathbf{x}^{(k+1)}) \leq \frac{\lambda_{\max(\mathbf{Q})} - \lambda_{\min(\mathbf{Q})}}{\lambda_{\max(\mathbf{Q})}} V(\mathbf{x}^{(k)}).$$

Definition 8.1. Given a sequence $\{\mathbf{x}^{(k)}\}$ that converges to $\mathbf{x}^*$, that is, $\lim_{k\to\infty} ||\mathbf{x}^{(k)} - \mathbf{x}^*|| = 0$, we say the order of convergence is p, where $p \in \mathbb{R}$, if

$$0 < \lim_{k\to\infty} \frac{||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||^p} < \infty.$$

If for all $p > 0$,

$$\lim_{k\to\infty} \frac{||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||^p} = 0,$$

then we say that the order of convergence is ∞.

Theorem 8.5. Let $\{\mathbf{x}^{(k)}\}$ be a sequence that converges to $\mathbf{x}^*$. If

$$||\mathbf{x}^{(k+1)} - \mathbf{x}^*|| = O(||\mathbf{x}^{(k)} - \mathbf{x}^*||^p),$$

then the order of convergence (if it exists) is at least p.

Lemma 8.3. In the steepest descent algorithm, if $\mathbf{g}^{(k)} \neq \mathbf{0}$ for all k, then $\gamma_k = 1$ if and only if $\mathbf{g}^{(k)}$ is an eigenvector of $\mathbf{Q}$.

3 Sufficient Condition for Local Minimizer for Problems with Linear Constraints

There are basically three types of sufficient conditions.

For non-constrained problems, if the point $\mathbf{x}^*$ is interior, then we need $\nabla f(\mathbf{x}^*) = 0$ and $\nabla f(\mathbf{x}^*) > 0$, i.e. positive definite.

For problems with linear equality constraints, i.e. minimize with respect to $\mathbf{A}\mathbf{x} = \mathbf{b}$, the sufficient condition says we need to satisfy the linear constraints. Also, consider a basis of the nullspace of A, then for every element $\mathbf{z}$ in this basis, we have $\mathbf{z}^T \nabla f(\mathbf{x}^*) = 0$ and $\mathbf{z}^T \nabla^2 f(\mathbf{x}^*)\mathbf{z} > 0$, i.e. positive definite.

For problems with linear inequality constraints, i.e. $\mathbf{A}\mathbf{x} \geq \mathbf{b}$, the sufficient condition says we need to satisfy the linear constraints. Also, $z^T \nabla f(\mathbf{x}^*) = 0$, and there exists λ^* such that $\lambda^* \geq 0$, $\lambda^* > 0$ if $A\mathbf{x} = 0$, and $A(\mathbf{x}^*)^T \lambda^* = 0$, and $\mathbf{z}^T \nabla^2 f(\mathbf{x}^*)\mathbf{z} > 0$, i.e. positive definite.

4 Newton Method

4.1 Recursive Formula

If $\mathbf{F}(\mathbf{x}^{(k)}) > 0$ (positive definite), then we have

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mathbf{F}(\mathbf{x}^{(k)})^{-1}\mathbf{g}^{(k)}.$$

4.2 Convergence

For quadratic function f, we have $\nabla f(\mathbf{x}^*) = \mathbf{0}$ after one step starting from any initial point $\mathbf{x}^{(0)}$.

Theorem 9.1. Suppose that $f \in \mathcal{C}^3$ and $\mathbf{x}^* \in \mathbb{R}^n$ is a point such that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ and $\mathbf{F}(\mathbf{x}^*)$ is invertible. Then for all $\mathbf{x}^{(0)}$ sufficiently close to $\mathbf{x}^*$, Newton's method is well-defined for all k and converges to $\mathbf{x}^*$ with an order of convergence at least 2.

Theorem 9.2. If we obtain a sequence $\{\mathbf{x}^{(k)}\}$ from Newton's method, and $\mathbf{F}(\mathbf{x}^{(k)}) > 0$ and $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$, then the search direction

$$\mathbf{d}^{(k)} = -\mathbf{F}(\mathbf{x}^{(k)})^{-1}\mathbf{g}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$$

is a descent direction for f in the sense that there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$,

$$f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) < f(\mathbf{x}^{(k)}).$$

4.3 Levenberg-Marquardt Modification

If the Hessian $\mathbf{F}(\mathbf{x}^{(k)})$ is not positive definite, then $\mathbf{d}^{(k)}$ may not point in a descent direction. Therefore, we modify it to be

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{F}(\mathbf{x}^{(k)}) + \mu_k \mathbf{I})^{-1}\mathbf{g}^{(k)},$$

where $\mu_k \geq 0$.

4.4 Newton's Method for Nonlinear Least Squares

Consider minimizing $\sum_{i=1}^{m}(r_i(\mathbf{x}))^2$ where $r_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \cdots, m$ are given functions. Define $\mathbf{r} = [r_1, \cdots, r_m]^T$, we can write $f(\mathbf{x}) = \mathbf{r}(\mathbf{x})^T\mathbf{r}(\mathbf{x})$. Denote the Jacobian matrix of $\mathbf{r}$ by $\mathbf{J}$, then the gradient is just

$$\nabla f(\mathbf{x}) = 2\mathbf{J}(\mathbf{x})^T\mathbf{r}(\mathbf{x}).$$

The Hessian of f is $\mathbf{F}(\mathbf{x}) = 2(\mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}) + \mathbf{S}(\mathbf{x}))$, where $\mathbf{S}(\mathbf{x})$ is the matrix with (k, j)th component as $\sum_{i=1}^{m} r_i(\mathbf{x})\frac{\partial^2 r_i}{\partial x_k \partial x_j}(\mathbf{x})$. Therefore the Newton's method applied here is given by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}) + \mathbf{S}(\mathbf{x}))^{-1}\mathbf{J}(\mathbf{x})^T\mathbf{r}(\mathbf{x}).$$

We can also ignore $\mathbf{S}$, so the Gauss-Newton method gives

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{J}(\mathbf{x})^T\mathbf{J}(\mathbf{x}))^{-1}\mathbf{J}(\mathbf{x})^T\mathbf{r}(\mathbf{x}).$$

5 Conjugate Gradient Method

5.1 Characteristic:

1. Solve quadratics of n variables in n Steps.
2. The usual implementation, the conjugate gradient algorithm, requires no Hessian matrix evaluation.
3. No matrix inversion and no storage of an $n \times n$ matrix are required.

Definition 10.1. Let $\mathbf{Q}$ be a real symmetric $n \times n$ matrix. The directions $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \cdots, \mathbf{d}^{(m)}$ are $\mathbf{Q}$-conjugate if for all $i \neq j$, we have $(\mathbf{d}^{(i)})^T \mathbf{Q} \mathbf{d}^{(j)} = 0$.

Lemma 10.1. Let $\mathbf{Q}$ be a symmetric positive definite $n \times n$ matrix. If the directions $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \cdots, \mathbf{d}^{(m)} \in \mathbb{R}^n$ with $k \leq n-1$ are nonzero and $\mathbf{Q}$-conjugate, then they are linearly independent.

5.2 Basic Conjugate Direction Algorithm

We aim to minimize a quadratic function of n variables $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{x}^T \mathbf{b}$, where $\mathbf{Q}$ is positive definite. Given a starting point $\mathbf{x}^{(0)}$ and $\mathbf{Q}$-conjugate directions $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \cdots, \mathbf{d}^{(m)}$; for $k \geq 0$,

$$\begin{aligned}
\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)}) &= \mathbf{Q}\mathbf{x}^{(k)} - \mathbf{b}, \\
\alpha_k &= -\frac{(\mathbf{g}^{(k)})^T \mathbf{d}^{(k)}}{(\mathbf{d}^{(k)})^T \mathbf{Q} \mathbf{d}^{(k)}}, \\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}.
\end{aligned}$$

Theorem 10.1. For any starting point $\mathbf{x}^{(0)}$, the basic conjugate direction algorithm converges to the unique $\mathbf{x}^*$ (that solves $\mathbf{Q}\mathbf{x} = \mathbf{b}$) in n steps.

Lemma 10.2. In the conjugate direction algorithm,

$$(\mathbf{g}^{(k+1)})^T \mathbf{d}^{(i)} = 0$$

for all k, $0 \leq k \leq n-1$ and $0 \leq i \leq k$.

5.3 Conjugate Gradient Algorithm

We aim to minimize a quadratic function of n variables $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{x}^T \mathbf{b}$, where $\mathbf{Q}$ is positive definite. Denote $\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k)})$.

1. Initially, take $\mathbf{x}^{(0)}$ and take $\mathbf{d}^{(0)} = -\mathbf{g}^{(0)}$.
2. Let $\alpha_k = -\frac{(\mathbf{g}^{(k)})^T \mathbf{d}^{(k)}}{(\mathbf{d}^{(k)})^T \mathbf{Q} \mathbf{d}^{(k)}}$.

3. Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$.
4. $\beta_k = \frac{(\mathbf{g}^{(k+1)})^T \mathbf{Q}\mathbf{d}^{(k)}}{(\mathbf{d}^{(k)})^T \mathbf{Q}\mathbf{d}^{(k)}}$.
5. $\mathbf{d}^{(k+1)} = -\mathbf{g}^{(k+1)} + \beta_k \mathbf{d}^{(k)}$. Loop back to take α_{k+1}.
6. Stop at $\mathbf{g}^{(k)} = 0$.

Proposition 10.1. In the conjugate gradient algorithm, the directions $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \cdots, \mathbf{d}^{(m)}$ are $\mathbf{Q}$-conjugate.

5.4 Conjugate Gradient Algorithm for Nonquadratic Problems

We interpret $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q}\mathbf{x} - \mathbf{x}^T \mathbf{b}$ as a second-order Taylor series approximation of the objective function.

5.5 Hestenes-Stiefel Formula

We update β_k by $\beta_k = \frac{(\mathbf{g}^{(k+1)})^T [\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}]}{(\mathbf{d}^{(k)})^T [\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}]}$.

5.6 Polak-Ribiere Formula

We update β_k by $\beta_k = \frac{(\mathbf{g}^{(k+1)})^T [\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}]}{(\mathbf{g}^{(k)})^T \mathbf{g}^{(k)}}$.

5.7 Fletcher-Reeves Formula

We update β_k by $\beta_k = \frac{(\mathbf{g}^{(k+1)})^T \mathbf{g}^{(k+1)}}{(\mathbf{g}^{(k)})^T \mathbf{g}^{(k)}}$.

6 Quasi-Newton Methods: BFGS Algorithm

In quasi-Newton methods, we want to approximate the Hessian matrix, instead of just calculating it and its inverse.

Therefore, quasi-Newton methods look like

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha H_k \mathbf{g}^{(k)}$$

where H_k is a matrix in place of the inverse Hessian.

6.1 How to approximate the inverse Hessian?

We have

$$\begin{aligned} \mathbf{d}^{(k)} &= -H_k \mathbf{g}^{(k)} \\ \alpha_k &= \arg\min_{\alpha \geq 0} f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \\ \mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}. \end{aligned}$$

where H_k is a symmetric matrix that acts as a continuous approximation of the Inverse Hessian.

In the quadratic case, we want it to satisfy

$$H_{k+1}\Delta\mathbf{g}^{(i)} = \Delta\mathbf{x}^{(i)}$$

for $0 \leq i \leq k$, where $\Delta\mathbf{x}^{(i)} = \alpha_k \mathbf{d}^{(k)}$.

6.2 BFGS Algorithm

Corresponding ot H_k, the algorithm updates B_k to be the approximation of the Hessian:

$$B_{k+1} = B_k + \frac{\Delta\mathbf{g}^{(k)}\Delta(\mathbf{g}^{(k)})^T}{\Delta(\mathbf{g}^{(k)})^T\Delta\mathbf{x}^{(k)}} - \frac{B_k\Delta\mathbf{x}^{(k)}\Delta(\mathbf{x}^{(k)})^T B_k}{\Delta(\mathbf{x}^{(k)})^T B_k \Delta\mathbf{x}^{(k)}}.$$

The approximation of the inverse Hessian is

$$H_{k+1}^{BFGS} = (B_{k+1})^{-1}.$$

6.3 Sherman-Morrison Formula

Note that $(A + uv^T)^{-1} = A^{-1} - \frac{(A^{-1}u)(v^T A^{-1})}{1+v^T A^{-1}u}$. Then we have another approximation given by

$$\begin{aligned} B_{k+1}^{BFGS} = B_k &+ \left(1 + \frac{\Delta(\mathbf{g}^{(k)})^T H_k \Delta\mathbf{g}^{(k)}}{\Delta(\mathbf{g}^{(k)})^T\Delta\mathbf{x}^{(k)}}\right)\frac{\Delta\mathbf{x}^{(k)}\Delta(\mathbf{x}^{(k)})^T}{\Delta(\mathbf{x}^{(k)})^T\Delta\mathbf{g}^{(k)}} \\ &- \frac{H_k\Delta\mathbf{g}^{(k)}\Delta(\mathbf{x}^{(k)})^T + (H_k\Delta\mathbf{g}^{(k)}\Delta(\mathbf{x}^{(k)})^T)^T}{\Delta(\mathbf{g}^{(k)})^T\Delta\mathbf{x}^{(k)}}. \end{aligned}$$

7 Linear Programming and Simplex Method

Formally, a linear program is an optimization problem of the form

$$\begin{aligned} \text{minimize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

7.1 Simplex Algorithm

1. Form a canonical augmented matrix corresponding to an initial basic feasible solution.
2. Calculate the reduced cost coefficients corresponding to the nonbasic variables.
3. If $r_j \geq 0$ for all j, stop – the current basic feasible solution is optimal.
4. Select a q such that $r_q < 0$.

5. If no $y_{iq} > 0$, stop – the problem is unbounded; else, calculate $p = \arg\min_i\{\frac{y_{i0}}{y_{iq}} : y_{iq} > 0\}$. (If more than one index i minimizes $\frac{y_{i0}}{y_{iq}}$, we let p be the smallest such index.)
6. Update the canonical augmented matrix by pivoting about the (p, q)-th element.
7. Go to step 2.

Example 16.2. Consider the linear program

$$\begin{aligned} \text{maximize } & 2x_1 + 5x_2 \\ \text{subject to } & x_1 \leq 4 \\ & x_2 \leq 6 \\ & x_1 + x_2 \leq 8 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Solve the problem by simplex method.

Proof. By introducing slack variables, we transform the problem into standard form and have

$$\begin{aligned} \text{minimize } & -2x_1 - 5x_2 \\ \text{subject to } & x_1 + x_3 = 4 \\ & x_2 + x_4 = 6 \\ & x_1 + x_2 + x_5 = 8 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0. \end{aligned}$$

Therefore, we should have the following canonical augmented matrix:

$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	$\mathbf{a_4}$	$\mathbf{a_5}$	$\mathbf{b}$
1	0	1	0	0	4
0	1	0	1	0	6
1	1	0	0	1	8
$\mathbf{-2}$	$\mathbf{-5}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	

Note that -5 is the smallest value (most negative), then we choose the pivot element from this column. We check the ratio of each element in $\mathbf{b}$ with corresponding element in $\mathbf{a}_2$, then we note that the smallest positive value is in row 2, which is 6. Therefore, the pivot element is $(2, 2)$. We now make the pivot element 1 (already is), and make other elements in this column 0 by row operations. This also includes the cost row (last row). Therefore, we have a new matrix

$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	$\mathbf{a_4}$	$\mathbf{a_5}$	$\mathbf{b}$
1	0	1	0	0	4
0	1	0	1	0	6
1	0	0	-1	1	2
$\mathbf{-2}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{5}$	$\mathbf{0}$	

Again, we note that the cost row still has a negative value, so we try to fix the issue in that column. The column is column 1, and the least positive value we need to fix in $\mathbf{a}_1$ is the element $(3, 1)$. We take this element as a pivot, and get

$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	$\mathbf{a_4}$	$\mathbf{a_5}$	$\mathbf{b}$
0	0	1	1	-1	2
0	1	0	1	0	6
1	0	0	-1	1	2
0	**0**	**0**	**3**	**2**	

This is the end because all entries in the cost row have non-negative values. Therefore, the current basic feasible solution is given by the last row, so $[2, 6, 2, 0, 0]^T$. In particular, we read the corresponding values and get $a_1 = 2$ and $a_2 = 6$ (from bottom to top because of the arrangement we have), and that is the optimal solution. Therefore, the function value we maximized is 34.

One remark is that if we fill in the bottom right value accordingly, the value at the end in that spot should be the value we want.

■

Example 16.3. Consider the linear program

$$\begin{aligned} \text{maximize } & 7x_1 + 6x_2 \\ \text{subject to } & 2x_1 + x_2 \leq 3 \\ & x_1 + 4x_2 \leq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Solve the problem by simplex method.

Proof. By introducing slack variables, we transform the problem into standard form and have

$$\begin{aligned} \text{minimize } & -7x_1 - 6x_2 \\ \text{subject to } & 2x_1 + x_2 + x_3 = 3 \\ & x_1 + 4x_2 + x_4 = 4 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

Therefore, we should have the following canonical augmented matrix:

$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	$\mathbf{a_4}$	$\mathbf{b}$
2	1	1	0	3
1	4	0	1	4
$\mathbf{-7}$	$\mathbf{-6}$	**0**	**0**	

Note that there are negative values in the cost row, so we focus on the most negative value here, which is in column 1. By procedure, we observe that the least positive ratio in this column is $\frac{3}{2} < 4$, so the pivot element is $(1,1)$. Therefore, we have

$$\begin{array}{ccccc}
\mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{b} \\
1 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{3}{2} \\
0 & \frac{7}{2} & -\frac{1}{2} & 1 & \frac{5}{2} \\
\mathbf{0} & -\mathbf{\frac{5}{2}} & \mathbf{\frac{7}{2}} & \mathbf{0} &
\end{array}$$

Note that there is still a negative value in the cost row, so we focus on the second column, where the pivot element should be $(2,2)$. Therefore, we have

$$\begin{array}{ccccc}
\mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{b} \\
1 & 0 & \frac{4}{7} & -\frac{1}{7} & \frac{8}{7} \\
0 & 1 & -\frac{1}{7} & \frac{2}{7} & \frac{5}{7} \\
\mathbf{0} & \mathbf{0} & \mathbf{\frac{22}{7}} & \mathbf{\frac{5}{7}} &
\end{array}$$

We can end here. Therefore, the basic feasible solution is given by $[\frac{8}{7}, \frac{5}{7}, 0, 0]^T$. In particular, the value of our function is $-8 - \frac{30}{7} = -\frac{86}{7}$. Therefore, the minimized value is $-\frac{86}{7}$, and so the maximized value of the original problem is $\frac{86}{7}$.

■

It is important to first standardize the matrix, i.e. get a basis from the canonical variables we have from the original problem. This is done automatically in Example 7.3. The following example requires more thinking.

Example (Exercise 17.3). Suppose we have the following problem:

$$\begin{aligned}
\text{minimize } & 4x_1 + 5x_2 \\
\text{subject to } & x_1 + 2x_2 \geq 2 \\
& 2x_1 + x_2 \geq 3 \\
& x_1, x_2 \geq 0.
\end{aligned}$$

We can standardize the problem with the following matrix:

$$\begin{array}{ccccc}
\mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{b} \\
1 & 2 & -1 & 0 & 2 \\
2 & 1 & 0 & -1 & 3 \\
4 & 5 & 0 & 0 & 0
\end{array}$$

Now we need to find a standard basis from $\mathbf{a}_1$ and $\mathbf{a}_2$. We want to take $(1,1)$ and $(2,2)$ as pivot, then we get the matrix

$$\begin{array}{ccccc}
\mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{b} \\
1 & 0 & \frac{1}{3} & -\frac{2}{3} & \frac{4}{3} \\
0 & 1 & -\frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 2 & 1 & -7
\end{array}$$

Note that the cost row at the bottom are all positive (ignore the last term). Therefore, we take $x_1 = \frac{4}{3}$ and $x_2 = \frac{1}{3}$, which gives us an optimal value 7.

8 Duality

Suppose that we are given a linear programming problem of the form

$$\begin{aligned} \text{minimize } & \mathbf{c}^T\mathbf{x} \\ \text{subject to } & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

We define the dual problem to be

$$\begin{aligned} \text{maximize } & \lambda^T\mathbf{b} \\ \text{subject to } & \lambda^T\mathbf{A} \leq \mathbf{c}^T \\ & \lambda \geq \mathbf{0}. \end{aligned}$$

Lemma 17.1. Suppose that $\mathbf{x}$ and λ are feasible solutions to primal and dual linear program problems, respectively (either in the symmetric or asymmetric form). Then, $\mathbf{c}^T\mathbf{x} \geq \lambda^T\mathbf{b}$.

Theorem 17.2. If the primal problem (either in symmetric or asymmetric form) has an optimal solution, then so does the dual, and the optimal values of their respective objective functions are equal.

Example 17.3. Consider the following linear programming problem:

$$\begin{aligned} \text{maximize } & 2x_1 + 5x_2 + x_3 \\ \text{subject to } & 2x_1 - x_2 + 7x_3 \leq 6 \\ & x_1 + 3x_2 + 4x_3 \leq 9 \\ & 3x_1 + 6x_2 + x_3 \leq 3 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

Then the dual problem to this is

$$\begin{aligned} \text{minimize } & 6x_1 + 9x_2 + 3x_3 \\ \text{subject to } & 2x_1 + x_2 + 3x_3 \geq 2 \\ & -x_1 + 3x_2 + 6x_3 \geq 5 \\ & 7x_1 + 4x_2 + x_3 \geq 1 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

Finally, we talk about the relation between the optimal solution between the primal problem and its corresponding dual problem. The solution to the dual is obtained by subtracting the reduced costs coefficients corresponding to the identity matrix of slack variables from the corresponding elements in the cost row vector.

Example 17.6. Consider example 17.3 above, find the optimal solution to the dual problem.

Proof. Recall that the standard form of the primal problem looks like

$$
\begin{array}{cccccccc}
 & \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \mathbf{a_5} & \mathbf{a_6} & \mathbf{b} \\
 & 2 & -1 & 7 & 1 & 0 & 0 & 6 \\
 & 1 & 3 & 4 & 0 & 1 & 0 & 9 \\
 & 3 & 6 & 1 & 0 & 0 & 1 & 3 \\
\mathbf{c}^T & -2 & -5 & -1 & 0 & 0 & 0 & 0
\end{array}
$$

By the simplex method, the final matrix should be

$$
\begin{array}{cccccccc}
 & \frac{15}{43} & 0 & 1 & \frac{6}{43} & 0 & \frac{1}{43} & \frac{39}{43} \\
 & -\frac{74}{43} & 0 & 0 & -\frac{21}{43} & 1 & -\frac{25}{43} & \frac{186}{43} \\
 & \frac{19}{43} & 1 & 0 & -\frac{1}{43} & 0 & \frac{7}{43} & \frac{15}{43} \\
\mathbf{r}^T & \frac{24}{43} & 0 & 0 & \frac{1}{43} & 0 & \frac{36}{43} & \frac{114}{43}
\end{array}
$$

The solution of the dual, $\lambda = [\lambda_1, \lambda_2, \lambda_3]$, can be solved by $\lambda^T \mathbf{D} = \mathbf{c}_D^T - \mathbf{r}_D^T$ (with respect to non-basic variables, so column 1, 4 and 6), so

$$[\lambda_1, \lambda_2, \lambda_3] \begin{pmatrix} 2 & 1 & 0 \\ 1 & 0 & 0 \\ 3 & 0 & 1 \end{pmatrix} = [-2, 0, 0] - [\frac{24}{43}, \frac{1}{43}, \frac{36}{43}].$$

Solving this equation yields the solution of the dual problem. ■

9 *Optional Topic*: Non-linear Constrained Optimization (Lagrangian, KKT methods)

Theorem 21.1 (KKT). Let $f, h, g \in \mathcal{C}^1$. Let $\mathbf{x}^*$ be a regular point and a local minimizer for the problem of minimizing f subject to $\mathbf{h}(x) = \mathbf{0}$, $\mathbf{g}(x) < \mathbf{0}$. Then, there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that

1. $\mu^* \geq 0$,
2. $Df(\mathbf{x}^*) + \lambda^{*T} D\mathbf{h}(\mathbf{x}^*) + \mu^{*T} D\mathbf{g}(\mathbf{x}^*) = \mathbf{0}^T$.
3. $\mu^{*T} g(\mathbf{x}^*) = 0$.

Here D is the gradient vector of the function, i.e. $D\mathbf{g}(\mathbf{x}^*) = \begin{pmatrix} \nabla g_1(\mathbf{x}^*)^T \\ \nabla g_2(\mathbf{x}^*)^T \end{pmatrix}$.