SWFR ENG 4TE3 (6TE3)

COMP SCI 4TE3 (6TE3)

Continuous Optimization Algorithm

# Constrained nonlinear optimization

Computing and Software
McMaster University

# Algorithms for constrained optimization

## Linear equality constraints

$$(LEC) \quad \begin{array}{ll} \text{min} & f(x) \\ \text{s.t.} & Ax = b. \end{array}$$

$f$ is continuously differentiable, $A : m \times n$ is a matrix with rank$(A) = m$ and $b \in \mathbb{R}^m$.

**Given a basis $B$ then**

$$Ax = Bx_B + Nx_N = b$$

and so we have

$$x_B = B^{-1}b - B^{-1}Nx_N.$$

We can rewrite (LEC) as

$$\text{min } f_N(x_N)$$

where $f_N(x_N) = f(x) = f(B^{-1}b - B^{-1}Nx_N; x_N)$.

This is an unconstrained problem. Further,

$$\nabla f(x)^T = ((\nabla_B f(x))^T, (\nabla_N f(x))^T),$$

The **reduced gradient** can be expressed as:

$$\begin{aligned} \nabla f_N(x_N)^T &= -(\nabla_B f(x))^T B^{-1}N + (\nabla_N f(x))^T \\ &= ((\nabla_B f(x))^T, (\nabla_N f(x))^T) \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}. \end{aligned}$$

The Reduced Hessian:

$$\nabla^2 f_N(x_N) = \begin{pmatrix} -(B^{-1}N)^T, & I \end{pmatrix} \nabla^2 f(x) \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}.$$

# Linear Equality Constraints

## Null-space method - I

$$(LEC) \quad \min \quad f(x)$$
$$\text{s.t.} \quad Ax = b.$$

$f$ is continuously differentiable, $A : m \times n$ is a matrix with rank$(A) = m$ and $b \in \mathbb{R}^m$.

Let $\bar{x}$ be feasible, i.e., $A\bar{x} = b$, then $Ax = A(\bar{x} + s) = b$, thus $(LEC)$ is equivalent to:

$$(LEC) \quad \min \quad f(\bar{x} + s)$$
$$\text{s.t.} \quad As = 0.$$

The vector $s$ is from the null-space of the matrix $A$.
If the columns of $Z$ (an $n \times (n-m)$ matrix) give a basis of the null-space of $A$, then $s = Zv$ with $v \in R^{n-m}$.
Then $(LEC)$ can be given by

$$(LEC) \qquad \min \quad h(v) = f(\bar{x} + Zv).$$

This is an unconstrained problem!

# Linear Equality Constraints

## Null-space method - II

**The null-space can easily be given:**
Let $B$ be a basis from the column space of $A$, then

$$A = (B,\ N)$$

The range(row) space of $A$ can be given by the basis vectors

$$R = (I,\ B^{-1}N).$$

The null-space is

$$Z^T = \left(-(B^{-1}N)^T,\ I\right).$$

Clearly $RZ = AZ = 0$.

The **gradient** of $h(v)$ can be expressed as

$$\nabla h(v) = Z^T \nabla f(\bar{x} + Zv).$$

The **Hessian** of $h(v)$ can be expressed as

$$\nabla^2 h(v) = Z^T \nabla^2 f(\bar{x} + Zv) Z.$$

# The reduced gradient method

## Linear (in)equality constraints

$$
\begin{array}{lll}
(LC) & \text{min} & f(x) \\
 & \text{s.t.} & Ax = b, \\
 & & x \geq 0.
\end{array}
$$

$f$ is continuously differentiable, $A : m \times n$ is a matrix with rank$(A) = m$ and $b \in \mathbb{R}^m$.

Given a basis $B$ and a feasible $x = (x_B, x_N)$ such that $x_B > 0$. $x_N$ do not have to be zero!

$$Bx_B + Nx_N = b$$

we have

$$x_B = B^{-1}b - B^{-1}Nx_N$$

$$
\begin{array}{ll}
\text{min} & f_N(x_N) \\
\text{s.t.} & B^{-1}b - B^{-1}Nx_N \geq 0, \\
 & x_N \geq 0,
\end{array}
$$

where $f_N(x_N) = f(x) = f(B^{-1}b - B^{-1}Nx_N, x_N)$ and

$$\nabla f(x)^T = ((\nabla_B f(x))^T, (\nabla_N f(x))^T).$$

The **reduced gradient** can be expressed as

$$r := \nabla f_N(x_N)^T = -(\nabla_B f(x))^T B^{-1} N + (\nabla_N f(x))^T.$$

# The reduced gradient method

## Linear constraints:

$x^k \in \mathbb{R}^n$ is the current iterate;
the basis is nondegenerate.
Search direction: $s^T = (s_B^T, s_N^T)$ in $\mathcal{N}(A)$
$s_B = -B^{-1}Ns_N$ and $s_N$ properly given,
the feasibility of $x^k + \lambda s$ is guaranteed as long as

$$x^k + \lambda s \geq 0, \quad \text{i.e.} \quad \lambda \leq \bar{\lambda} = \min_{1 \leq i \leq n,\ s_i < 0} \left\{ \frac{x_i^k}{-s_i} \right\}.$$

Further, $s_N$ should be a descent direction of $f$.

$$s_j = \begin{cases} 0 & \text{if } \ x_j^k = 0 \ \text{ and } r_j \geq 0, \\ -r_j & \text{otherwise} \end{cases} \qquad j \in N.$$

Make a line search: $x^{k+1} = \arg\min_{0 \leq \lambda \leq \bar{\lambda}} f(x^k + \lambda s)$.

If all the coordinates $x_B^{k+1}$ stay strictly positive we keep the basis, else a pivot is made to eliminate the zero variable from the basis and replace it by a positive but currently non-basic coordinate.

# The reduced gradient method

## Convergent variant:

$x^k \in \mathbb{R}^n$ is the current iterate;
the basis is nondegenerate.
Search direction: $s^T = (s_B^T, s_N^T)$ in $\mathcal{N}(A)$
$s_B = -B^{-1}Ns_N$ and $s_N$ is given by

$$s_j = \begin{cases} -x_j r_j & \text{if } \ r_j \geq 0, \\ -r_j & \text{otherwise} \end{cases} \quad j \in N.$$

The feasibility of $x^k + \lambda s$ is guaranteed as long as

$$x^k + \lambda s \geq 0, \quad \text{i.e.} \quad \lambda \leq \bar{\lambda} = \min_{1 \leq i \leq n,\, s_i < 0} \left\{ \frac{x_i^k}{-s_i} \right\}.$$

**Theorem 1.** *The search direction $s$ at $x^k$ is always a descent direction unless $s = 0$. If $s = 0$, then $x^k$ is a KKT point of problem (LC).*

**Theorem 2.** *Any accumulation point of the sequence $\{x^k\}$ is a KKT point.*

# SQP-I

## Sequential quadratic programming
## Equality constraints

$$
\begin{array}{lll}
(NC) & \text{min} & f(x) \\
 & \text{s.t.} & h_j(x) = 0, \quad j = 1, \cdots, m
\end{array}
$$

The Lagrange function is

$$L(x, y) = f(x) + \sum_{j=1}^{m} y_j h_j(x),$$

where $y_j \in R, \quad j = 1, \cdots, m$. Let us denote

$$H(x) = (h_1(x), \cdots, h_m(x))^T.$$

Then the KKT conditions are:

$$
\begin{array}{rcl}
\nabla_x L(x, y) & = & 0 \\
H(x) & = & 0.
\end{array}
$$

Let a candidate solution $(x^k, y^k)$ be given and apply Newton's method to solve this nonlinear equation system:

$$
\begin{array}{rcrcl}
\nabla_{xx}^2 L(x^k, y^k)\Delta x & + & (\nabla H(x^k))^T \Delta y & = & -\nabla_x L(x^k, y^k) \\
\nabla H(x^k)\Delta x & & & = & -H(x^k).
\end{array}
$$

# SQP-II

**Sequential quadratic programming**
**Equality constraints**

This equation system

$$\begin{array}{rcl} \nabla^2_{xx} L(x^k, y^k)\Delta x \ + (\nabla H(x^k))^T \Delta y & = & -\nabla_x L(x^k, y^k) \\ \nabla H(x^k)\Delta x & = & -H(x^k) \end{array}$$

is the KKT condition of the following linearly constrained quadratic optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\Delta x^T \nabla^2_{xx} L(x^k, y^k)\Delta x + \nabla_x L(x^k, y^k)^T \Delta x \\ \text{s.t.} & \nabla H(x^k)\Delta x = -H(x^k). \end{array}$$

One needs to:

– solve this quadratic problem at each iteration,

– make a line-search where feasibility and optimality need to be considered

– and repeat the process from the new point until the optimality condition is satisfied.

# Barrier functions

How can one replace the constraint $t \geq 0$ (i.e., $-g_j(x) \geq 0$) by a good barrier function?

Desired properties of barrier function $B(t)$ of $t \geq 0$:

1. $B(t)$ is a smooth (infinitely many times) differentiable, strictly convex.
2. The derivative of $B(t)$ goes to $-\infty$ as $t \to 0$.
3. $B(t)$ goes to infinity as $t \to 0$.

**Note:**
For barrier functions you need inequality constraints!

# Barrier functions

Examples:

1. The *logarithmic barrier* function $-\log t$.
2. Let $r > 1$. The *inverse barrier* function $t^{-r}$.

---

The barrier function for the $(CO)$ problem

$$
\begin{array}{lll}
(CO) & \text{min} & f(x) \\
 & \text{s.t.} & g_j(x) \leq 0, \;\; j = 1, \cdots, m
\end{array}
$$

is given by

$$f_\mu(x) = \frac{f(x)}{\mu} + \sum_{j=1}^{m} B(-g_j(x)) \tag{1}$$

where $\mu > 0$. The original problem (CO) is solved by sequentially minimizing the function $f_\mu(x)$ for a series of $\mu$ values as $\mu \to 0$.

# Generic Interior Point Method

**Input**:

$\mu = \mu_0$ the barrier parameter value;

$\theta$ the reduction parameter, $0 < \theta < 1$;

$\epsilon > 0$ the accuracy parameter;

$x^0$ a given interior feasible point;

**Step 0:** $x := x^0$, $\mu := \mu_0$;

**Step 1:** If $\mu < \epsilon$ STOP, $x(\mu)$ is returned as the solution.

**Step 2:** Calculate (approximately) $x(\mu)$;

**Step 3:** $\mu := (1 - \theta)\mu$;

**Step 4:** GO TO Step 1.

# Log-barrier methods

## and Lagrange multiplier estimates

The log-barrier function for the $(CO)$ problem

$$
\begin{array}{lll}
(CO) & \min & f(x) \\
 & \text{s.t.} & g_j(x) \le 0, \quad j = 1, \ldots, m
\end{array}
$$

for $\mu > 0$ is given by $\quad f_\mu(x) = f(x) + \mu \sum_{j=1}^m - \log(-g_j(x))$.

The optimality condition when minimizing $f_\mu(x)$ is:

$$
\nabla f(x) + \sum_{j=1}^m \frac{\mu}{-g_j(x)} \nabla g_j(x) = 0. \qquad (*)
$$

On the other hand, the Lagrange function for $(CO)$ is

$$
L(x, y) = f(x) + \sum_{j=1}^m y_j g_j(x).
$$

In the Wolfe dual we get the constraint:

$$
\nabla f(x) + \sum_{j=1}^m y_j \nabla g_j(x) = 0, \qquad (**)
$$

the optimality condition to minimize $L(x, y)$ in $x$.
Comparing $(*)$ and $(**)$ we have that

$$
\frac{\mu}{-g_j(x)} \qquad \text{is an estimate of} \qquad y_j,
$$

the Lagrange multiplier.

# Penalty functions

How can one force
the equality constraints $t = 0$ (i.e., $h_i(x) = 0$)) and
the inequality constraints $t \leq 0$ (i.e, $g_j(x) \leq 0$)
by penalizing the non-satisfaction of these constraints?
What are the desirable properties of a penalty function?

Desired properties of a penalty function $P(t)$:

1. $P(t)$ is nonnegative and strictly convex;
2. $P(t) = 0$ for feasible points;
3. $P(t)$ goes to infinity as infeasibility increases;
4. $P(t)$ increases sharply as infeasibility occurs;
5. $P(t)$ is a smooth (infinitely many times) differentiable.

Functions satisfying at least the first three properties are called *penalty functions.*

# Penalty functions

**For the equality constraints $t = 0$ (i.e., $h_i(x) = 0$))**

- Quadratic penalty function: $P(t) = t^2$.
- Exact penalty function: $P(t) = |t|$.

**For the inequality constraints $t \le 0$ (i.e, $g_j(x) \le 0$)**

1. Quadratic penalty function:
   $P(t) = \begin{cases} 0 & \text{if } t \le 0; \\ t^2 & \text{if } t > 0; \end{cases} = (\max\{0, t\})^2.$
2. Exact penalty function: $P(t) = \max\{0, t\}$.

---

$$
\begin{array}{lll}
(CO) & \min & f(x) \\
 & \text{s.t.} & g_j(x) \le 0, \ \ j = 1, \cdots, m, \\
 & & h_i(x) = 0, \ \ i = 1, \cdots, k.
\end{array}
$$

The quadratic penalty function for $(CO)$ is:

$$
P(x) = f(x) + \vartheta \left( \sum_{j=1}^{m} (\max\{0, g_j(x)\})^2 + \sum_{i=1}^{k} (h_i(x))^2 \right).
$$

The exact penalty function for $(CO)$ is:

$$
P(x) = f(x) + \vartheta \left( \sum_{j=1}^{m} \max\{0, g_j(x)\} + \sum_{i=1}^{k} |h_i(x)| \right).
$$