Algorithms for nonlinear programming problems II

Martin Branda

Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Computational Aspects of Optimization

Algorithm convergence

Definition

Let $X \subseteq \mathbb{R}^p$, $Y \subseteq \mathbb{R}^q$ be nonempty closed sets. Let $F : X \to Y$ be a set-valued mapping. The map F is said to be **closed** at $x \in X$ if for any sequences $\{x^k\} \subset X$, and $\{y^k\}$ satisfying $x_k \to x$, $y^k \in F(x^k)$, $y^k \to y$ we have that $y \in F(x)$.
The map F is said to be closed on $Z \subseteq X$ if it is closed at each point in Z.

Algorithm convergence – Zangwill's theorem

Bazaraa et al. (2006), Theorem 7.2.3: Let

A1. $X \subseteq \mathbb{R}^p$ be a nonempty **closed** set

A2. $\hat{X} \subseteq X$ be a **nonempty solution set**

A3. $F : X \to X$ be a set-valued mapping **closed** over complement of $\hat{X}$

A4. Given $x^1 \in X$ the sequence $\{x^k\}$ is generated iteratively as follows: If $x^k \in \hat{X}$, then STOP; otherwise, let $x^{k+1} \in F(x^k)$ and repeat.

A5. the sequence $x^1, x^2, \ldots$ be contained in a **compact** subset of X

A6. there exist a **continuous function**[1] α such that $\alpha(y) < \alpha(x)$ if $x \notin \hat{X}$ and $y \in F(x)$

Then either the algorithm stops in a finite number of steps with a point in $\hat{X}$ or it generates an infinite sequence $\{x^k\}$ such that all accumulation points belong to $\hat{X}$ and $\alpha(x^k) \to \alpha(x)$ for some $x \in \hat{X}$.

[1] descent function: $\alpha(x) = f(x)$ or $\alpha(x) = \|\nabla f(x)\|$

Algorithm convergence – Newton method

Let $\overline{x}$ be an optimal solution, set

$$\begin{aligned} F(x) &= x - \left(\nabla^2 f(x)\right)^{-1} \nabla f(x), \\ \alpha(x) &= \|x - \overline{x}\|. \end{aligned}$$

More details: Bazaraa et al. (2006), Theorem 8.6.5

Algorithm classification

- **Order of derivatives**[2]**:** derivative-free, first order (gradient), second-order (Newton)
- **Feasibility of the constructed points**: interior and exterior point methods
- **Deterministic/randomized**
- **Local**/global

[2] If possible, deliver the derivatives.

Barrier method

$f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$

$$\min_x f(x) \text{ s.t. } g_j(x) \le 0,\ j = 1, \dots, m.$$

(Extension including equality constraints is straightforward.)

Assumption

$$\{x : \ g(x) < 0\} \neq \emptyset.$$

Barrier method

Barrier functions: continuous with the following properties

$$B(y) \geq 0, y < 0, \lim_{y \to 0_-} B(y) = \infty.$$

Examples

$$\frac{-1}{y}, \; -\log \min\{1, -y\}.$$

Maybe the most popular barrier function:

$$B(y) = -\log -y$$

Set

$$\tilde{B}(x) = \sum_{j=1}^{m} B\big(g_j(x)\big),$$

and solve

$$\min f(x) + \mu \tilde{B}(x), \tag{1}$$

where $\mu > 0$ is a parameter.

Polynomial-time interior point methods for LP

$$\min\ c^T x - \mu \sum_{j=1}^{n} \log x_j \qquad \text{s.t. } Ax = b. \tag{2}$$

KKT-conditions ...

Interior point methods

$f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$

$$\min_x f(x) \text{ s.t. } g_j(x) \le 0,\ j = 1, \dots, m.$$

(Extension including equality constraints is straightforward.)

New slack decision variables $s \in \mathbb{R}^m$

$$\begin{aligned} \min_{x,s}\ & f(x) \\ \text{s.t. } & g_j(x) + s_j = 0,\ j = 1, \dots, m, \\ & s_j \ge 0. \end{aligned} \tag{3}$$

Interior point methods

Barrier problem

$$\begin{aligned} \min_{x,s}\ & f(x) - \mu \sum_{j=1}^{m} \log s_j \\ \text{s.t. } & g(x) + s = 0. \end{aligned} \tag{4}$$

The barrier term prevents the components of s from becoming too close to zero.

Lagrangian function

$$L(x,s,z) = f(x) - \mu \sum_{j=1}^{m} \log s_j - \sum_{j=1}^{m} z_j (g_j(x) + s_j).$$

Interior point methods

KKT conditions for barrier problem (matrix notation)

$$\begin{aligned}
\nabla f(x) - \nabla g^T(x) z &= 0,\\
-\mu S^{-1} e - z &= 0,\\
g(x) + s &= 0,
\end{aligned}$$

$S = \mathrm{diag}\{s_1, \ldots, s_m\}$, $Z = \mathrm{diag}\{z_1, \ldots, z_m\}$, $\nabla g(x)$ is the Jacobian matrix (components of function in rows?)

Multiply the second equality by S

$$\begin{aligned}
\nabla f(x) - \nabla g^T(x) z &= 0,\\
-SZe &= \mu e,\\
g(x) + s &= 0,
\end{aligned}$$

$=$ Nonlinear system of equalities $\rightarrow$ Newton's method

Newton's method

$$\nabla f(x+v) = \nabla f(x) + \nabla^2 f(x) v = 0$$

with the solution (under $\nabla^2 f(x) \succ 0$)

$$v = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$$

Interior point methods

Use Newton's method to obtain a step $(\Delta x, \Delta s, \Delta z)$

$$\begin{pmatrix} H(x,z) & 0 & -\nabla g^T \\ 0 & -Z & -S \\ \nabla g & I & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta z \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + \nabla g^T(x) z \\ SZe + \mu e \\ -g(x) - s \end{pmatrix}$$

$H(x,z) = \nabla^2 f(x) - \sum_{j=1}^m z_j \nabla^2 g_j(x)$, $\nabla^2 f$ denotes the Hessian matrix.

Interior point methods

Stopping criterion:

$$E = \max \left\{ \left\| \nabla f(x) - \nabla g^T(x) z \right\|, \| SZe + \mu e \|, \| g(x) + s \| \right\} \leq \varepsilon,$$

$\varepsilon > 0$ small.

Interior point methods

ALGORITHM:

0. Choose x^0 and $s^0 > 0$, and compute initial values for the multipliers $z^0 > 0$. Select an initial barrier parameter $\mu^0 > 0$ and parameter $\sigma \in (0, 1)$, set $k = 1$.
1. Repeat until a stopping test for the nonlinear program (19.1) is satisfied:
    - Solve the nonlinear system of equalities using Newton's method and obtain (x^k, s^k, z^k).
    - Decrease barrier parameter $\mu^{k+1} = \sigma\mu^k$, set $k = k + 1$.

Interior point methods

Convergence of the method (Nocedal and Wright 2006, Theorem 19.1): continuously differentiable f, g_j, LICQ at any limit point, then the limits are stationary points of the original problem

Interior point method – Example

$$\begin{aligned} \min_{x}\ & (x_1-2)^4+(x_1-2x_2)^2 \\ \text{s.t. } & x_1^2-x_2\le 0. \end{aligned} \tag{5}$$

$$\begin{aligned} \min_{x,s}\ & (x_1-2)^4+(x_1-2x_2)^2 \\ \text{s.t. } & x_1^2-x_2+s=0, \\ & s\ge 0. \end{aligned} \tag{6}$$

$$\begin{aligned} \min_{x,s}\ & (x_1-2)^4+(x_1-2x_2)^2-\mu\log s \\ \text{s.t. } & x_1^2-x_2+s=0. \end{aligned} \tag{7}$$

Interior point method – Example

Lagrange function

$$L(x_1, x_2, s, z) = (x_1 - 2)^4 + (x_1 - 2x_2)^2 - \mu \log s - z(x_1^2 - x_2 + s).$$

Optimality conditions together with feasibility

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 4(x_1 - 2)^3 + 2(x_1 - 2x_2) - 2zx_1 = 0,\\
\frac{\partial L}{\partial x_2} &= -4(x_1 - 2x_2) + z = 0,\\
\frac{\partial L}{\partial s} &= -\frac{\mu}{s} - z = 0,\\
\frac{\partial L}{\partial z} &= x_1^2 - x_2 + s = 0.
\end{aligned} \tag{8}$$

We have obtained 4 equations with 4 variables ...

Interior point method – Example

Slight modification

$$\begin{aligned}
4(x_1-2)^3+2(x_1-2x_2)-2zx_1 &= 0, && (9)\\
-4(x_1-2x_2)+z &= 0, && (10)\\
-sz-\mu &= 0, && (11)\\
x_1^2-x_2+s &= 0. && (12)
\end{aligned}$$

Necessary derivatives

$$\begin{aligned}
H(x_1,x_2,z) &= \begin{pmatrix} 12(x_1-2)^2+2-2z & -4 \\ -4 & 8 \end{pmatrix} && (13)\\
\nabla g(x) &= \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix} && (14)
\end{aligned}$$

Interior point method – Example

System of linear equations for Newton's step

$$\begin{pmatrix} 12(x_1-2)^2+2-2z & -4 & 0 & -2x_1 \\ -4 & 8 & 0 & 1 \\ 0 & 0 & -z & -s \\ 2x_1 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta s \\ \Delta z \end{pmatrix} =$$

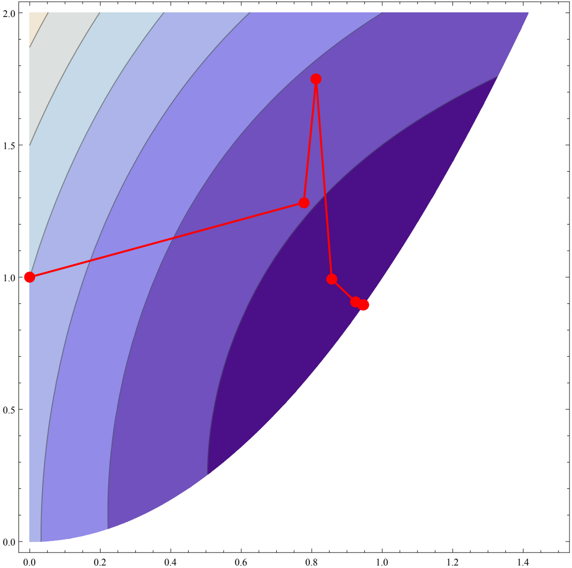
$$= \begin{pmatrix} -4(x_1-2)^3 - 2(x_1-2x_2) + 2zx_1 \\ 4(x_1-2x_2) - z \\ sz + \mu \\ -x_1^2 + x_2 - s \end{pmatrix}$$

Interior point method – Example

Starting point $x^0 = (0, 1)$, $z^0 = 1$, $s^0 = 1$, $\mu > 0$, then the step ...

$$\begin{pmatrix} 48 & -4 & 0 & 0 \\ -4 & 8 & 0 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta s \\ \Delta z \end{pmatrix} = \begin{pmatrix} 36 \\ -9 \\ 1+\mu \\ 0 \end{pmatrix}$$

Interior point method – Example



Sequential Quadratic Programming

Nocedal and Wright (2006): Let $f, h_i : \mathbb{R}^n \to \mathbb{R}$ be smooth functions,

$$\begin{aligned} &\min_x \; f(x) \\ &\text{s.t. } h_i(x) = 0, \;\; i = 1, \dots, l. \end{aligned} \tag{15}$$

Lagrange function

$$L(x, v) = f(x) + \sum_{i=1}^{l} v_i h_i(x)$$

and KKT optimality conditions

$$\begin{aligned} \nabla_x L(x, v) &= \nabla_x f(x) + A(x)^T v = 0, \\ h(x) &= 0, \end{aligned} \tag{16}$$

where $h(x)^T = (h_1(x), \dots, h_l(x))$ and
$A(x)^T = [\nabla h_1(x), \nabla h_2(x), \dots, \nabla h_l(x)]$ denotes the Jacobian matrix.

Sequential Quadratic Programming

We have system of $n + l$ equations in the $n + l$ unknowns x and v:

$$\nabla F(x, v) = \begin{bmatrix} \nabla_x f(x) + A(x)^T v \\ h(x) \end{bmatrix} = 0. \tag{17}$$

ASS. (LICQ) Jacobian matrix $A(x)$ has full row rank.
The Jacobian is given by

$$\nabla^2 F(x, v) = \begin{bmatrix} \nabla^2_{xx} L(x, v) & A(x)^T \\ A(x) & 0 \end{bmatrix}. \tag{18}$$

We can use the **Newton algorithm** ..

Sequential Quadratic Programming

Setting $f_k = f(x^k)$, $\nabla^2_{xx} L_k = \nabla^2_{xx} L(x^k, v^k)$, $A_k = A(x^k)$, $h_k = h(x^k)$, we obtain the Newton step by solving the system

$$\begin{bmatrix} \nabla^2_{xx} L_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} p_x \\ p_v \end{bmatrix} = \begin{bmatrix} -\nabla f_k - A_k^T v_k \\ -h_k \end{bmatrix} \tag{19}$$

Then we set $x^{k+1} = x^k + p_x$ and $v^{k+1} = v^k + p_v$.

Sequential Quadratic Programming

ASS. (SOSC) For all $d \in \{\tilde{d} \neq 0 : \; A(x)\tilde{d} = 0\}$, it holds

$$d^T \nabla^2_{xx} L(x, v) d > 0.$$

Sequential Quadratic Programming

Important alternative way to see the Newton iterations: Consider the **quadratic program**

$$\begin{aligned} \min_p\ & f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2_{xx} L_k p \\ \text{s.t. } & h_k + A_k p = 0. \end{aligned} \tag{20}$$

KKT optimality conditions

$$\begin{aligned} \nabla^2_{xx} L_k p + \nabla f_k + A_k^T \tilde{v} &= 0 \\ h_k + A_k p &= 0, \end{aligned} \tag{21}$$

$$\begin{bmatrix} \nabla^2_{xx} L_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} p_x \\ p_{\tilde{v}} \end{bmatrix} = \begin{bmatrix} -\nabla f_k \\ -h_k \end{bmatrix} \tag{22}$$

which is the same as the Newton system if we add $A_k^T v_k$ to the first equation. Then we set $x^{k+1} = x^k + p_x$ and $v^{k+1} = p_{\tilde{v}}$.

Sequential Quadratic Programming

Algorithm: Start with an initial solution (x^0, v^0) and iterate until a convergence criterion is met:

1. Evaluate $f_k = f(x^k)$, $h_k = h(x^k)$, $A_k = A(x^k)$, $\nabla^2_{xx} L_k = \nabla^2_{xx} L(x^k, v^k)$.
2. Solve the Newton equations OR the quadratic problem to obtain new (x^{k+1}, v^{k+1}).

If possible, deliver explicit formulas for first and second order derivatives.

Sequential Quadratic Programming

$f, g_j, h_i : \mathbb{R}^n \to \mathbb{R}$ are smooth. We use a **quadratic approximation** of the objective function and linearize the constraints, $p \in \mathbb{R}^p$

$$\begin{aligned} \min_p \ & f(x^k) + p^T \nabla_x f(x^k) + \frac{1}{2} p^T \nabla^2_{xx} L(x^k, u^k, v^k) p \\ \text{s.t. } & g_j(x^k) + p^T \nabla_x g_j(x^k) \le 0, \ j = 1, \ldots, m, \\ & h_i(x^k) + p^T \nabla_x h_i(x^k) = 0, \ i = 1, \ldots, l. \end{aligned} \tag{23}$$

Use an algorithm for quadratic programming to solve the problem and set $x^{k+1} = x^k + p_k$, where u^{k+1}, v^{k+1} are Lagrange multipliers of the quadratic problem which are used to compute new $\nabla^2_{xx} L$.

Convergence: Nocedal and Wright (2006), Theorem 18.1

Literature

- Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. (2006). **Nonlinear programming: theory and algorithms**, Wiley, Singapore, 3rd edition.
- Boyd, S., Vandenberghe, L. (2004). **Convex Optimization**. Cambridge University Press, Cambridge.
- P. Kall, J. Mayer: **Stochastic Linear Programming: Models, Theory, and Computation**. Springer, 2005.
- Nocedal, J., Wright, J.S. (2006). **Numerical optimization**. Springer, New York, 2nd edition.