Numerical Optimization

Unconstrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Newton Method

Consider the problem,

$$\min_{\boldsymbol{x}} \ f(\boldsymbol{x})$$

- Let $f \in \mathcal{C}^2$ and f be bounded below.
- Newton method uses quadratic approximation of f at a given point, $\boldsymbol{x}^k$

$$f(\boldsymbol{x}) \approx f_q^k(\boldsymbol{x}) = f(\boldsymbol{x}^k) + \boldsymbol{g}^{k^T}(\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^k)^T \boldsymbol{H}^k (\boldsymbol{x} - \boldsymbol{x}^k)$$

- $\boldsymbol{x}^{k+1}$ is the minimizer of $f_q^k(\boldsymbol{x})$
- What is the *region of trust* in which this approximation is reliable?

Trust-Region Method

- Given $\Delta^k > 0$, let the *region of trust* be Ω^k where

$$\Omega^k = \{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{x}^k\| \leq \Delta^k\}$$

- Solve the following constrained problem to get $\boldsymbol{x}^{k+1}$:

$$\begin{array}{ll} \min & f_q^k(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \Omega^k \end{array}$$

- How to determine Ω^{k+1} (or Δ^{k+1})? Can use the *actual* and *predicted* reduction in f

Trust-Region Method

Algorithm to determine Δ^{k+1} and R^k

(1) Given $\Delta^k, \boldsymbol{x}^k, \boldsymbol{x}^{k+1}$

(2) $R^k = \frac{f(\boldsymbol{x}^k) - f(\boldsymbol{x}^{k+1})}{f_q^k(\boldsymbol{x}^k) - f_q^k(\boldsymbol{x}^{k+1})}$

(3) **if** $R^k < 0.25$

$$\Delta^{k+1} = \|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|/4$$

else if $R^k > 0.75$ **and** $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| == \Delta^k$

$$\Delta^{k+1} = 2\Delta^k$$

else

$$\Delta^{k+1} = \Delta^k$$

endif

Output : Δ^{k+1}, R^k

Modified Newton Algorithm (based on Trust Region)

(1) Initialize $\boldsymbol{x}^0$, ϵ and Δ^0, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x} \in \Omega^k} f_q^k(\boldsymbol{x})$

  (b) Determine Δ^{k+1}, R^k

  (c) If $R^k < 0$, $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k$

  (d) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

Quasi-Newton Methods

Consider the problem,

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

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$.

- Let $f \in \mathcal{C}^2$.
  - Newton method:

    $$f(\boldsymbol{x}) \approx f_q^k(\boldsymbol{x}) = f(\boldsymbol{x}^k) + \boldsymbol{g}^{k^T}(\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^k)^T \boldsymbol{H}^k (\boldsymbol{x} - \boldsymbol{x}^k)$$

  - Newton direction: $\boldsymbol{d}_N^k = -(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k$
- Given $f \in \mathcal{C}^1$, form a quadratic model of f at $\boldsymbol{x}^k$:

  $$y_k(\boldsymbol{x}) = f(\boldsymbol{x}^k) + \boldsymbol{g}^{k^T}(\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^k)^T \boldsymbol{B}^{k^{-1}} (\boldsymbol{x} - \boldsymbol{x}^k)$$

  where $\boldsymbol{B}^k$ is a symmetric positive definite matrix.
- Quasi-Newton direction: $\boldsymbol{d}_{QN}^k = -\boldsymbol{B}^k\boldsymbol{g}^k$

Quasi-Newton Methods

$$y_k(\boldsymbol{x}) = f(\boldsymbol{x}^k) + {\boldsymbol{g}^k}^T(\boldsymbol{x} - \boldsymbol{x}^k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^k)^T {\boldsymbol{B}^k}^{-1}(\boldsymbol{x} - \boldsymbol{x}^k)$$

- $(\boldsymbol{B}^k)^{-1}$ is either $\boldsymbol{H}^k$ or its approximation
- $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k_{QN} = \boldsymbol{x}^k - \alpha^k \boldsymbol{B}^k \boldsymbol{g}^k$
- Given $\boldsymbol{x}^k, \boldsymbol{x}^{k+1}, \boldsymbol{g}^k, \boldsymbol{g}^{k+1}$ and $\boldsymbol{B}^k$, how to update $\boldsymbol{B}^k$ to get a symmetric positive definite matrix $\boldsymbol{B}^{k+1}$?
- Is $\boldsymbol{B}^k \approx (\boldsymbol{H}^k)^{-1}$?
- Are there any conditions that $\boldsymbol{B}^{k+1}$ should satisfy?

Given $\boldsymbol{x}^{k+1}$, we construct a quadratic approximation of f at $\boldsymbol{x}^{k+1}$:

$$y_{k+1}(\boldsymbol{x}) = f(\boldsymbol{x}^{k+1}) + {\boldsymbol{g}^{k+1}}^T(\boldsymbol{x}-\boldsymbol{x}^{k+1}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{k+1})^T(\boldsymbol{B}^{k+1})^{-1}(\boldsymbol{x}-\boldsymbol{x}^{k+1})$$

Require

$$\begin{aligned} \nabla y_{k+1}(\boldsymbol{x}^k) &= \nabla f(\boldsymbol{x}^k) \\ \nabla y_{k+1}(\boldsymbol{x}^{k+1}) &= \nabla f(\boldsymbol{x}^{k+1}) = \boldsymbol{g}^{k+1} \end{aligned}$$

Therefore, we require,

$$\nabla y_{k+1}(\boldsymbol{x}^k) = \nabla f(\boldsymbol{x}^k) = \boldsymbol{g}^k = \boldsymbol{g}^{k+1} + (\boldsymbol{B}^{k+1})^{-1}(\boldsymbol{x}^k - \boldsymbol{x}^{k+1})$$

Letting $\boldsymbol{g}^{k+1} - \boldsymbol{g}^k = \boldsymbol{\gamma}^k$ and $\boldsymbol{x}^{k+1} - \boldsymbol{x}^k = \boldsymbol{\delta}^k$, we get

$$\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^k = \boldsymbol{\delta}^k \quad \text{(Quasi-Newton condition)}$$

- Quasi-Newton condition

$$\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^k = \boldsymbol{\delta}^k$$

- $\boldsymbol{B}^{k+1}$ should be positive definite

$$\boldsymbol{\gamma}^{k^T}\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^k = \boldsymbol{\gamma}^{k^T}\boldsymbol{\delta}^k > 0 \ \ \forall\, \boldsymbol{\gamma}^k \neq 0$$

- From Wolfe conditions for line search,

$$\boldsymbol{g}^{k+1^T}\boldsymbol{d}^k \geq c_2\, \boldsymbol{g}^{k^T}\boldsymbol{d}^k,\ c_2 \in (0,1) \ \Rightarrow\ \boldsymbol{\gamma}^{k^T}\boldsymbol{\delta}^k > 0$$

  $\therefore$ When Wolfe condition is satisfied in a line search, $\exists\, \boldsymbol{B}^{k+1}$ which satisfies Quasi-Newton condition
- $\frac{n(n+1)}{2}$ variables to be found using n equations and n inequalities

Consider a simple way to update $\boldsymbol{B}^k$: Let $\alpha \neq 0, \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u} \neq \mathbf{0}$

$$\boldsymbol{B}^{k+1} = \boldsymbol{B}^k + \alpha \boldsymbol{u}\boldsymbol{u}^T \qquad \text{(Rank-one correction)}$$

Choose α and $\boldsymbol{u}$ such that $\boldsymbol{B}^{k+1}$ satisfies *Quasi-Newton condition*

$$\begin{aligned} \therefore (\boldsymbol{B}^k + \alpha \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{\gamma}^k &= \boldsymbol{\delta}^k \\ \therefore \alpha \boldsymbol{u}^T \boldsymbol{\gamma}^k \boldsymbol{u} &= \boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k \end{aligned}$$

Let $\boldsymbol{u} = \boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k$.
Therefore, $\alpha \boldsymbol{u}^T \boldsymbol{\gamma}^k = 1$ gives $\alpha^{-1} = \left(\boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k\right)^T \boldsymbol{\gamma}^k$.

$$\therefore \boldsymbol{B}_{SR1}^{k+1} = \boldsymbol{B}^k + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k\right)\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k\right)^T}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k \boldsymbol{\gamma}^k\right)^T \boldsymbol{\gamma}^k}$$

$\boldsymbol{B}^{k+1}$ obtained using $\boldsymbol{x}^k, \boldsymbol{x}^{k+1}, \boldsymbol{g}^k$ and $\boldsymbol{g}^{k+1}$

Quasi-Newton Algorithm (rank-one correction)

(1) Initialize $\boldsymbol{x}^0$, ϵ and symmetric positive definite $\boldsymbol{B}^0$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\boldsymbol{d}^k = -\boldsymbol{B}^k\boldsymbol{g}^k$

  (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that

    (i) $f(\boldsymbol{x}^k + \alpha^k\boldsymbol{d}^k) < f(\boldsymbol{x}^k)$

    (ii) α^k satisfies Armijo-Wolfe (or Armijo-Goldstein) conditions

  (c) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k\boldsymbol{d}^k$

  (d) Find $\boldsymbol{B}^{k+1}$ using rank-one correction

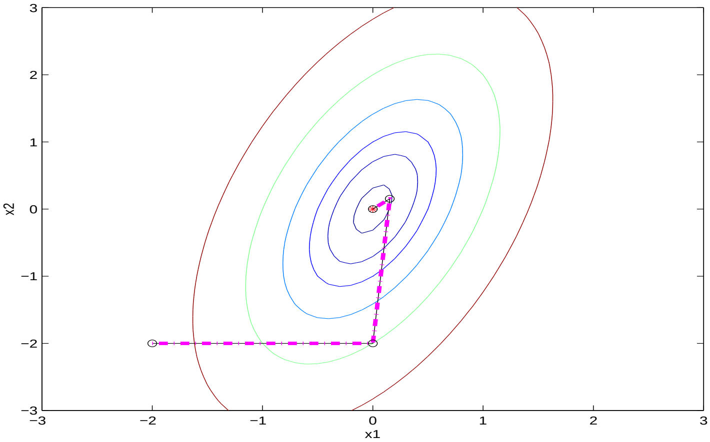
  (e) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

Example:

$$\min \quad f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Quasi-Newton algorithm (rank-one correction) with inexact line search applied to $f(\boldsymbol{x})$

Example:

$$\min \ f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

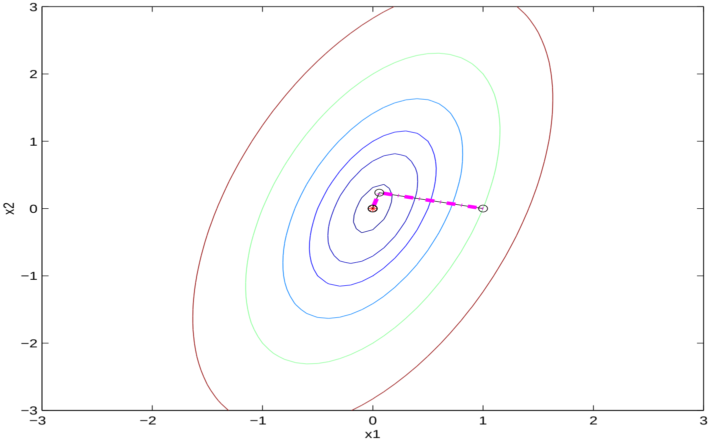
- $\boldsymbol{x}^* = (0,0)^T, \boldsymbol{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix},$

  $\boldsymbol{H}^{-1} = \begin{pmatrix} 0.1667 & 0.1667 \\ 0.1667 & 0.6667 \end{pmatrix}$

k	x_1^k	x_2^k	B^k		$\|\|\boldsymbol{g}^k\|\|$
0	-2	-2	1	0	12.0
			0	1	
1	0	-2	0.1833	0.2333	5.65
			0.2333	0.9333	
2	.1538	.1536	0.1667	0.1667	0.92
			0.1667	0.6667	
3	0	0			0

Example:

$$\min \quad f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Quasi-Newton algorithm (rank-one correction) with inexact line search applied to $f(\boldsymbol{x})$

Example:

$$\min \; f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

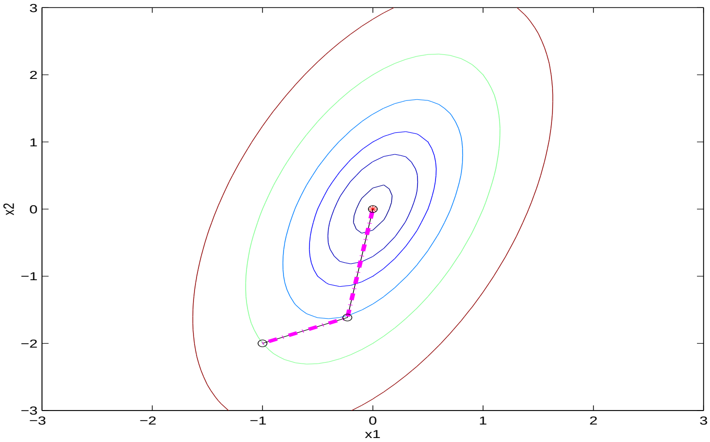
- $\boldsymbol{x}^* = (0,0)^T, \boldsymbol{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix},$

  $\boldsymbol{H}^{-1} = \begin{pmatrix} 0.1667 & 0.1667 \\ 0.1667 & 0.6667 \end{pmatrix}$

k	x_1^k	x_2^k	B^k		$\Vert\boldsymbol{g}^k\Vert$
0	1	0	1	0	8.25
			0	1	
1	0.0588	.2353	0.1892	0.2432	0.35
			0.2432	0.9270	
2	-0.0029	0	0.1667	0.1667	0.024
			0.1667	0.6667	
3	0	0			0

Example:

$$\min \quad f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Quasi-Newton algorithm (rank-one correction) with inexact line search applied to $f(\boldsymbol{x})$

Example:

$$\min \; f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

- $\boldsymbol{x}^* = (0,0)^T, \boldsymbol{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$,
  $\boldsymbol{H}^{-1} = \begin{pmatrix} 0.1667 & 0.1667 \\ 0.1667 & 0.6667 \end{pmatrix}$

k	x_1^k	x_2^k	B^k		$\lVert \boldsymbol{g}^k \rVert$
0	-1	-2	1	0	4.47
			0	1	
1	-0.2308	-1.6154	0.1724	0.2069	3.09
			0.2069	0.9483	
2	0	0			0

Consider the problem,

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}$$

where $\boldsymbol{H}$ is a symmetric positive definite matrix.
Newton Method: Choose any $\boldsymbol{x}^0, \boldsymbol{d}_N^0 = -\boldsymbol{H}^{-1}\boldsymbol{g}^0, \boldsymbol{x}_1 = \boldsymbol{x}^*$.

- Suppose we apply *Quasi-Newton* method (rank-one correction) to solve this problem
- At every iteration k,
  - $\boldsymbol{B}^{k+1}$ is symmetric positive definite
  - $\boldsymbol{B}^{k+1}$ is obtained from $\boldsymbol{B}^k, \boldsymbol{x}^k, \boldsymbol{x}^{k+1}, \boldsymbol{g}^k$ and $\boldsymbol{g}^{k+1}$
  - $\boldsymbol{B}^{k+1}$ satisfies Quasi-Newton condition, $\boldsymbol{B}^{k+1}\gamma^k = \delta^k$

  Note that,

$$\begin{aligned} \boldsymbol{g}^k &= \boldsymbol{H}\boldsymbol{x}^k + \boldsymbol{c} \\ \boldsymbol{g}^{k+1} &= \boldsymbol{H}\boldsymbol{x}^{k+1} + \boldsymbol{c} \end{aligned}$$

$$\therefore \boldsymbol{g}^{k+1} - \boldsymbol{g}^k = \boldsymbol{H}(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) \;\Rightarrow\; \gamma^k = \boldsymbol{H}\delta^k$$

Using Quasi-Newton condition at every iteration, we have

$$
\begin{aligned}
k = 0, &\quad \boldsymbol{B}^1 \boldsymbol{\gamma}^0 = \boldsymbol{\delta}^0 \\
k = 1, &\quad \boldsymbol{B}^2 \boldsymbol{\gamma}^1 = \boldsymbol{\delta}^1 \\
k = 2, &\quad \boldsymbol{B}^3 \boldsymbol{\gamma}^2 = \boldsymbol{\delta}^2 \\
&\qquad\quad \vdots \quad \vdots \\
k = n-1, &\quad \boldsymbol{B}^n \boldsymbol{\gamma}^{n-1} = \boldsymbol{\delta}^{n-1}
\end{aligned}
$$

In addition to Quasi-Newton condition at every iteration, if we ensure

$$
\begin{aligned}
k = 0, \boldsymbol{B}^1\gamma^0 &= \boldsymbol{\delta}^0 \\
k = 1, \boldsymbol{B}^2\gamma^1 = \boldsymbol{\delta}^1, \boldsymbol{B}^2\gamma^0 &= \boldsymbol{\delta}^0 \\
k = 2, \quad \boldsymbol{B}^3\gamma^2 = \boldsymbol{\delta}^2, \boldsymbol{B}^3\gamma^1 = \boldsymbol{\delta}^1, \boldsymbol{B}^3\gamma^0 &= \boldsymbol{\delta}^0 \\
&\vdots \quad \vdots \\
k = n-1, \quad \boldsymbol{B}^n\gamma^{n-1} = \boldsymbol{\delta}^{n-1}, \boldsymbol{B}^n\gamma^{n-2} = \boldsymbol{\delta}^{n-2}, \ldots \boldsymbol{B}^n\gamma^0 &= \boldsymbol{\delta}^0
\end{aligned}
$$

Hereditary Property

$$k = n - 1, \quad \boldsymbol{B}^n\boldsymbol{\gamma}^{n-1} = \boldsymbol{\delta}^{n-1}, \boldsymbol{B}^n\boldsymbol{\gamma}^{n-2} = \boldsymbol{\delta}^{n-2}, \ldots \boldsymbol{B}^n\boldsymbol{\gamma}^{0} = \boldsymbol{\delta}^{0}$$

Suppose $(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k)^T\boldsymbol{\gamma}^k \neq 0$ in the rank-one correction.

$$\therefore \boldsymbol{B}^n \left(\gamma^{n-1}|\ldots|\gamma^1|\gamma^0\right) = \left(\boldsymbol{\delta}^{n-1}|\ldots|\boldsymbol{\delta}^1|\boldsymbol{\delta}^0\right)$$

Using $\gamma^k = \boldsymbol{H}\boldsymbol{\delta}^k$ for every k, we have

$$\boldsymbol{B}^n\boldsymbol{H} \left(\boldsymbol{\delta}^{n-1}|\ldots|\boldsymbol{\delta}^1|\boldsymbol{\delta}^0\right) = \left(\boldsymbol{\delta}^{n-1}|\ldots|\boldsymbol{\delta}^1|\boldsymbol{\delta}^0\right)$$

If $\boldsymbol{\delta}^0, \boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{n-1}$ *are linearly independent*, then

$$\boldsymbol{B}^n\boldsymbol{H} = \boldsymbol{I} \Rightarrow \boldsymbol{B}^n = \boldsymbol{H}^{-1}$$

Therefore, after n iterations, $\boldsymbol{d}^n_{QN} = -\boldsymbol{B}^n\boldsymbol{g}^n = -\boldsymbol{H}^{-1}\boldsymbol{g}^n = \boldsymbol{d}^n_N$ and

$$\boldsymbol{x}^{n+1} = \boldsymbol{x}^*.$$

For a convex quadratic function, the solution is attained in at most $n + 1$ iterations using rank-one correction for $\boldsymbol{B}^k$.

Hereditary Property

For the symmetric rank-one correction applied to a quadratic function with positive definite Hessian $\boldsymbol{H}$,

$$\boldsymbol{B}^k\boldsymbol{\gamma}^j = \boldsymbol{\delta}^j, \ \ j = 0, \ldots, k-1$$

Proof.

Note that $\boldsymbol{H}\boldsymbol{\delta}^k = \boldsymbol{\gamma}^k \ \forall \ k$.
For $k = 1, \boldsymbol{B}^1\boldsymbol{\gamma}^0 = \boldsymbol{\delta}^0$. (Quasi-Newton condition)
Suppose $\boldsymbol{B}^k\boldsymbol{\gamma}^j = \boldsymbol{\delta}^j, j = 0, \ldots, k-1$.
Using rank-one correction and using $j = 0, \ldots, k-1$,

$$\begin{aligned} \boldsymbol{B}^{k+1} &= \boldsymbol{B}^k + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k} \\ \therefore \boldsymbol{B}^{k+1}\boldsymbol{\gamma}^j &= \left(\boldsymbol{B}^k + \frac{(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k)((\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k)^T}{(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k)^T\boldsymbol{\gamma}^k}\right)\boldsymbol{\gamma}^j \end{aligned}$$

Proof. (continued)

$$
\begin{aligned}
\therefore \boldsymbol{B}^{k+1}\boldsymbol{\gamma}^j &= \boldsymbol{B}^k\boldsymbol{\gamma}^j + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k}\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^j \\
&= \boldsymbol{B}^k\boldsymbol{\gamma}^j + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k}\left(\boldsymbol{\delta}^{kT}\boldsymbol{\gamma}^j - \boldsymbol{\gamma}^{kT}\boldsymbol{B}^k\boldsymbol{\gamma}^j\right) \\
&= \boldsymbol{B}^k\boldsymbol{\gamma}^j + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k}\left(\boldsymbol{\delta}^{kT}\boldsymbol{H}\boldsymbol{\delta}^j - \boldsymbol{\delta}^{kT}\boldsymbol{H}\boldsymbol{\delta}^j\right) \\
&= \boldsymbol{B}^k\boldsymbol{\gamma}^j \\
&= \boldsymbol{\delta}^j \ \forall\, j = 0, \ldots, k-1
\end{aligned}
$$

Also, $\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^k = \boldsymbol{\delta}^k$ (Quasi-Newton condition)

Therefore,

$$\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^j = \boldsymbol{\delta}^j \ \forall\, j = 0, \ldots, k$$

$\square$

Theorem

Consider the problem,

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}$$

where $\boldsymbol{H}$ *is a symmetric positive definite matrix.*
If the rank-one correction is well defined and $\boldsymbol{\delta}^0, \boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{n-1}$ *are linearly independent, then the rank-one correction method applied to minimize* $f(\boldsymbol{x})$ *terminates in at most* $n+1$ *iterations, with* $\boldsymbol{B}^n = \boldsymbol{H}^{-1}$.

Quasi-Newton Methods - Rank One correction

$$\boldsymbol{B}^{k+1} = \boldsymbol{B}^k + \frac{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T}{\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k}$$

Some Remarks:

- A simple and elegant way to use the information gathered during two consecutive iterations to update $\boldsymbol{B}^k$
- $\boldsymbol{B}^{k+1}$ is *positive definite* if $\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k > 0$ *which cannot be guaranteed at every k*
- Numerical difficulties if $\left(\boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k\right)^T\boldsymbol{\gamma}^k \approx 0$

The following update methods have received wide acceptance:

- Davidon-Fletcher-Powell (DFP) method
- Broyden-Fletcher-Goldfarb-Shanno (BFGS) method

Rank Two Correction

Given that $\boldsymbol{B}^k$ is symmetric and positive definite matrix, let

$$\boldsymbol{B}^{k+1} = \boldsymbol{B}^k + \alpha \boldsymbol{u}\boldsymbol{u}^T + \beta \boldsymbol{v}\boldsymbol{v}^T$$

Quasi-Newton condition, $\boldsymbol{B}^{k+1}\boldsymbol{\gamma}^k = \boldsymbol{\delta}^k$ gives

$$\alpha \boldsymbol{u}^T\boldsymbol{\gamma}^k \boldsymbol{u} + \beta \boldsymbol{v}^T\boldsymbol{\gamma}^k \boldsymbol{v} = \boldsymbol{\delta}^k - \boldsymbol{B}^k\boldsymbol{\gamma}^k$$

Letting $\alpha \boldsymbol{u}^T\boldsymbol{\gamma}^k = 1$ and $\beta \boldsymbol{v}^T\boldsymbol{\gamma}^k = 1$ gives

$$\alpha^{-1} = \boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k$$
$$\beta^{-1} = -\boldsymbol{\gamma}^{k^T}\boldsymbol{B}^k\boldsymbol{\gamma}^k$$

Therefore,

$$\boldsymbol{B}^{k+1} = \boldsymbol{B}^k + \frac{\boldsymbol{\delta}^k\boldsymbol{\delta}^{k^T}}{\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k} - \frac{\boldsymbol{B}^k\boldsymbol{\gamma}^k\boldsymbol{\gamma}^{k^T}\boldsymbol{B}^k}{\boldsymbol{\gamma}^{k^T}\boldsymbol{B}^k\boldsymbol{\gamma}^k} \quad \text{(DFP Method)}$$

$$\boldsymbol{B}_{DFP}^{k+1} = \boldsymbol{B}^k + \frac{\boldsymbol{\delta}^k \boldsymbol{\delta}^{k^T}}{\boldsymbol{\delta}^{k^T} \boldsymbol{\gamma}^k} - \frac{\boldsymbol{B}^k \boldsymbol{\gamma}^k \boldsymbol{\gamma}^{k^T} \boldsymbol{B}^k}{\boldsymbol{\gamma}^{k^T} \boldsymbol{B}^k \boldsymbol{\gamma}^k} \quad \text{(DFP Method)}$$

- Is $\boldsymbol{B}_{DFP}^{k+1}$ a symmetric positive definite matrix, given that $\boldsymbol{B}^k$ is symmetric positive definite matrix?

$\boldsymbol{B}_{DFP}^{k+1}$ is a symmetric matrix.
Let $\boldsymbol{x} \neq \mathbf{0}, \boldsymbol{\gamma}^k \neq \mathbf{0}, \boldsymbol{\delta}^k \neq \mathbf{0}$.

$$\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{B}^k \boldsymbol{x} - \frac{(\boldsymbol{x}^T \boldsymbol{B}^k \boldsymbol{\gamma}^k)^2}{\boldsymbol{\gamma}^{k^T} \boldsymbol{B}^k \boldsymbol{\gamma}^k} + \frac{(\boldsymbol{\delta}^{k^T} \boldsymbol{x})^2}{\boldsymbol{\delta}^{k^T} \boldsymbol{\gamma}^k}$$

Since $\boldsymbol{B}^k$ is symmetric, $\boldsymbol{B}^k = \boldsymbol{B}^{k^{\frac{1}{2}}} \boldsymbol{B}^{k^{\frac{1}{2}}}$ where $\boldsymbol{B}^{k^{\frac{1}{2}}}$ is symmetric and positive definite.
Letting $\boldsymbol{a} = \boldsymbol{B}^{k^{\frac{1}{2}}} \boldsymbol{x}$ and $\boldsymbol{b} = \boldsymbol{B}^{k^{\frac{1}{2}}} \boldsymbol{\gamma}^k$,

$$\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} = \frac{(\boldsymbol{a}^T \boldsymbol{a})(\boldsymbol{b}^T \boldsymbol{b}) - (\boldsymbol{a}^T \boldsymbol{b})^2}{\boldsymbol{b}^T \boldsymbol{b}} + \frac{(\boldsymbol{\delta}^{k^T} \boldsymbol{x})^2}{\boldsymbol{\delta}^{k^T} \boldsymbol{\gamma}^k}$$

$$\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} = \frac{(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) - (\boldsymbol{a}^T\boldsymbol{b})^2}{\boldsymbol{b}^T\boldsymbol{b}} + \frac{(\boldsymbol{\delta}^{k^T}\boldsymbol{x})^2}{\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k}$$

- $(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) \geq (\boldsymbol{a}^T\boldsymbol{b})^2$ (Cauchy-Schwartz inequality)
- $\boldsymbol{b}^T\boldsymbol{b} = \boldsymbol{\gamma}^{k^T}\boldsymbol{B}^k\boldsymbol{\gamma}^k > 0$ ($\boldsymbol{B}^k$ is a positive definite matrix)
  Note that $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha^k\boldsymbol{B}^k\boldsymbol{g}^k \Rightarrow \boldsymbol{\delta}^k = -\alpha^k\boldsymbol{B}^k\boldsymbol{g}^k$
  Suppose that $\boldsymbol{x}^{k+1}$ is obtained using exact line search.
  $\therefore \boldsymbol{g}^{k+1^T}\boldsymbol{\delta}^k = 0$
-
$$\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k = \boldsymbol{\delta}^{k^T}(\boldsymbol{g}^{k+1} - \boldsymbol{g}^k) = -\boldsymbol{g}^k\boldsymbol{\delta}^k = \alpha^k\boldsymbol{g}^{k^T}\boldsymbol{B}^k\boldsymbol{g}^k > 0$$

Therefore, $\boldsymbol{x}^T\boldsymbol{B}_{DFP}^{k+1}\boldsymbol{x} \geq 0$, or $\boldsymbol{B}_{DFP}^{k+1}$ is positive semi-definite.

$$\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} = \frac{(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) - (\boldsymbol{a}^T\boldsymbol{b})^2}{\boldsymbol{b}^T\boldsymbol{b}} + \frac{(\boldsymbol{\delta}^{k^T}\boldsymbol{x})^2}{\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k}$$

We now show that $\boldsymbol{B}_{DFP}^{k+1}$ is positive definite, that is,

$$\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} > 0, \boldsymbol{x} \neq \boldsymbol{0}$$

We have already shown that $\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k > 0$.
Suppose $\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} = 0, \boldsymbol{x} \neq \boldsymbol{0}$.
Therefore, $(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) = (\boldsymbol{a}^T\boldsymbol{b})^2$ and $(\boldsymbol{\delta}^{k^T}\boldsymbol{x})^2 = 0$.

$$(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{b}^T\boldsymbol{b}) = (\boldsymbol{a}^T\boldsymbol{b})^2 \;\Rightarrow\; \boldsymbol{a} = \mu\boldsymbol{b} \;\Rightarrow\; \boldsymbol{x} = \mu\boldsymbol{\gamma}^k \;\Rightarrow\; \mu \neq 0$$

$$(\boldsymbol{\delta}^{k^T}\boldsymbol{x})^2 = 0 \;\Rightarrow\; \mu\boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k = 0 \;\Rightarrow\; \boldsymbol{\delta}^{k^T}\boldsymbol{\gamma}^k = 0 \;\; \text{(contradiction)}$$

Therefore, $\boldsymbol{x}^T \boldsymbol{B}_{DFP}^{k+1} \boldsymbol{x} > 0, \boldsymbol{x} \neq \boldsymbol{0} \;\Rightarrow\; \boldsymbol{B}_{DFP}^{k+1}$ is positive definite.

Quasi-Newton Algorithm (DFP Method)

(1) Initialize $\boldsymbol{x}^0$, ϵ and symmetric positive definite $\boldsymbol{B}^0$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\boldsymbol{d}^k = -\boldsymbol{B}^k \boldsymbol{g}^k$

  (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that

    (i) $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) < f(\boldsymbol{x}^k)$

    (ii) α^k satisfies Armijo-Wolfe (or Armijo-Goldstein) conditions

  (c) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

  (d) Find $\boldsymbol{B}^{k+1}$ using DFP method

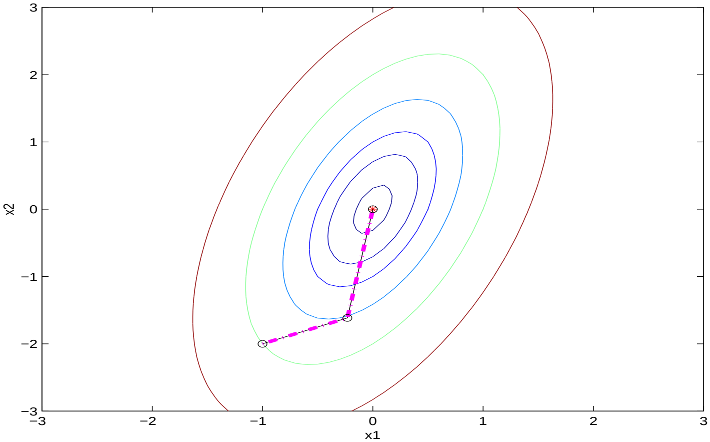
  (e) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

Example:

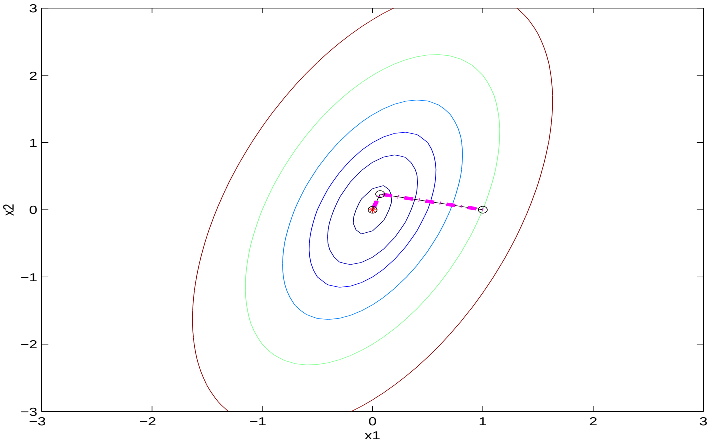
$$\min \quad f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2 - 2x_1x_2$$



Quasi-Newton algorithm with exact line search applied to $f(\boldsymbol{x})$

Example:

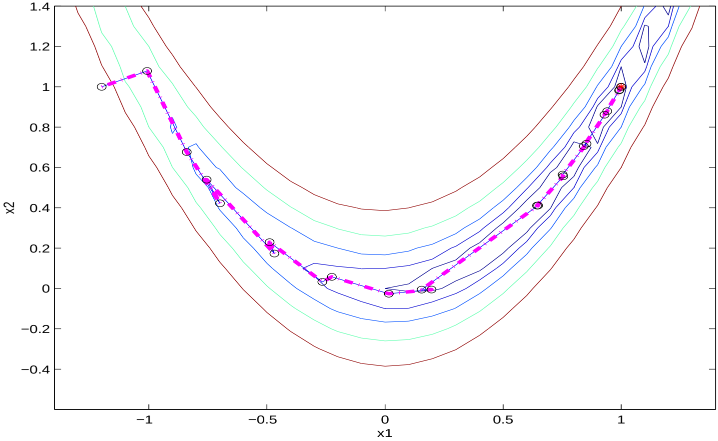
$$\min \quad f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$



Quasi-Newton algorithm with exact line search applied to $f(\boldsymbol{x})$

Example (Rosenbrock function):

$$\min \ f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of Quasi-Newton algorithm (DFP Method) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (-1.2, 1)^T$

Example (Rosenbrock function):

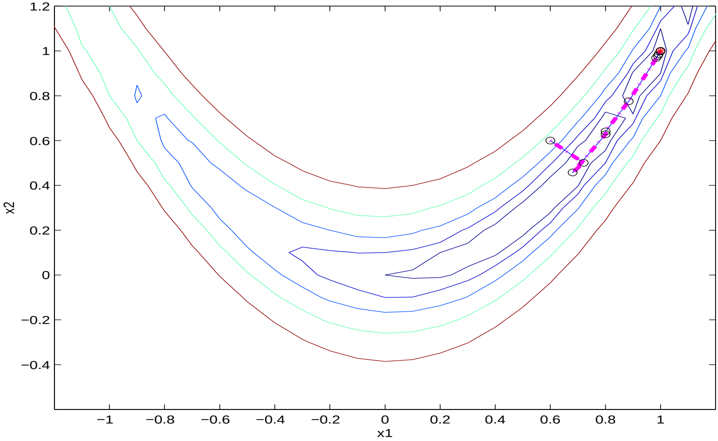
$$\min \quad f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

k	x_1^k	x_2^k	$f(\boldsymbol{x}^k)$	$\lVert\boldsymbol{g}^k\rVert$	$\lVert\boldsymbol{x}^k - \boldsymbol{x}^*\rVert$
0	-1.2	1	24.2	232.86	2.2
1	-1.01	1.08	4.43	24.97	2.01
2	-0.84	0.68	3.47	14.53	1.87
3	-0.70	0.42	3.33	25.61	1.79
4	-0.76	0.54	3.18	14.19	1.81
5	-0.47	0.17	2.37	14.80	1.69
10	0.20	-0.01	0.84	9.00	1.29
15	0.75	0.56	0.06	0.34	0.50
20	0.99	0.99	0.0002	0.69	0.02
24	0.99	0.99	5.72×10^{-12}	2.25×10^{-6}	5.35×10^{-6}

Table: Quasi-Newton algorithm (DFP Method) applied to Rosenbrock function, using $\boldsymbol{x}^0 = (-1.2, 1.0)^T$.

Example (Rosenbrock function):

$$\min \; f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of Quasi-Newton algorithm (DFP Method) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (0.6, 0.6)^T$

Example (Rosenbrock function):

$$\min \; f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

k	x_1^k	x_2^k	$f(\boldsymbol{x}^k)$	$\|\boldsymbol{g}^k\|$	$\|\boldsymbol{x}^k - \boldsymbol{x}^*\|$
0	0.6	0.6	5.92	75.60	0.57
1	0.72	0.50	0.12	6.32	0.572095
2	0.68	0.46	0.11	1.56	0.629112
3	0.80	0.63	0.06	4.65	0.421985
4	0.80	0.64	0.04	0.39	0.410591
5	0.88	0.78	0.02	2.67	0.252278
6	0.99	0.98	0.0005	0.94	0.0238278
7	0.98	0.97	0.0003	0.33	0.0348487
8	0.99	0.99	7.8×10^{-5}	0.23	0.02
9	0.99	0.99	5.3×10^{-7}	0.0048	0.0016
10	0.99	0.99	1.1×10^{-8}	0.0044	3.2×10^{-5}
11	0.99	0.99	3.1×10^{-13}	3.2×10^{-6}	1.2×10^{-6}

Table: Quasi-Newton algorithm (DFP Method) applied to Rosenbrock function, using $\boldsymbol{x}^0 = (0.6, 0.6)^T$.

- Newton direction, $\boldsymbol{d}_N^k = -(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k$
- Quasi-Newton direction, $\boldsymbol{d}_{QN}^k = -\boldsymbol{B}^k\boldsymbol{g}^k$
    - Quasi-Newton condition: $\boldsymbol{B}^{k+1}\gamma^k = \boldsymbol{\delta}^k$
      $\Rightarrow\ \gamma^k = (\boldsymbol{B}^{k+1})^{-1}\boldsymbol{\delta}^k$
- Let $\boldsymbol{G}^{k+1} = (\boldsymbol{B}^{k+1})^{-1}$ approximate $\boldsymbol{H}^{k+1}$.
  Therefore, we get *dual* formulae:

$$\boldsymbol{B}^{k+1}\gamma^k = \boldsymbol{\delta}^k$$
$$\boldsymbol{G}^{k+1}\boldsymbol{\delta}^k = \gamma^k$$

-

$$\boldsymbol{B}_{DFP}^{k+1} = \boldsymbol{B}^k + \frac{\boldsymbol{\delta}^k{\boldsymbol{\delta}^k}^T}{{\boldsymbol{\delta}^k}^T\gamma^k} - \frac{\boldsymbol{B}^k\gamma^k{\gamma^k}^T\boldsymbol{B}^k}{{\gamma^k}^T\boldsymbol{B}^k\gamma^k} \quad \text{(DFP Method)}$$

$$\boldsymbol{G}_{BFGS}^{k+1} = \boldsymbol{G}^k + \frac{\gamma^k{\gamma^k}^T}{{\gamma^k}^T\boldsymbol{\delta}^k} - \frac{\boldsymbol{G}^k\boldsymbol{\delta}^k{\boldsymbol{\delta}^k}^T\boldsymbol{G}^k}{{\boldsymbol{\delta}^k}^T\boldsymbol{G}^k\boldsymbol{\delta}^k} \quad \text{(BFGS Method)}$$

$$\boldsymbol{G}_{BFGS}^{k+1} = \boldsymbol{G}^k + \frac{\boldsymbol{\gamma}^k \boldsymbol{\gamma}^{k^T}}{\boldsymbol{\gamma}^{k^T} \boldsymbol{\delta}^k} - \frac{\boldsymbol{G}^k \boldsymbol{\delta}^k \boldsymbol{\delta}^{k^T} \boldsymbol{G}^k}{\boldsymbol{\delta}^{k^T} \boldsymbol{G}^k \boldsymbol{\delta}^k} \quad \text{(BFGS Method)}$$

- How to get $\boldsymbol{B}_{BFGS}^{k+1}$ from $\boldsymbol{G}_{BFGS}^{k+1}$?
- Use the condition,

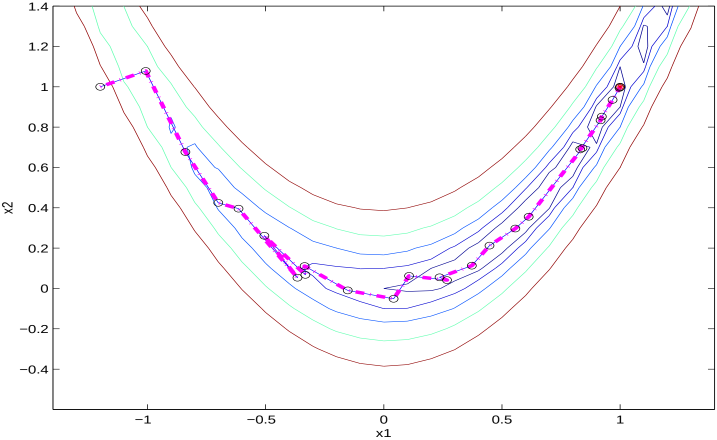
$$\boldsymbol{B}_{BFGS}^{k+1} \boldsymbol{G}_{BFGS}^{k+1} = \boldsymbol{I}$$

to get

$$\boldsymbol{B}_{BFGS}^{k+1} = \boldsymbol{B} + \left(1 + \frac{\boldsymbol{\gamma}^T \boldsymbol{B} \boldsymbol{\gamma}}{\boldsymbol{\delta}^T \boldsymbol{\gamma}}\right) \frac{\boldsymbol{\delta} \boldsymbol{\delta}^T}{\boldsymbol{\delta}^T \boldsymbol{\gamma}} - \left(\frac{\boldsymbol{\delta} \boldsymbol{\gamma}^T \boldsymbol{B} + \boldsymbol{B} \boldsymbol{\gamma} \boldsymbol{\delta}^T}{\boldsymbol{\delta}^T \boldsymbol{\gamma}}\right)$$

Example (Rosenbrock function):

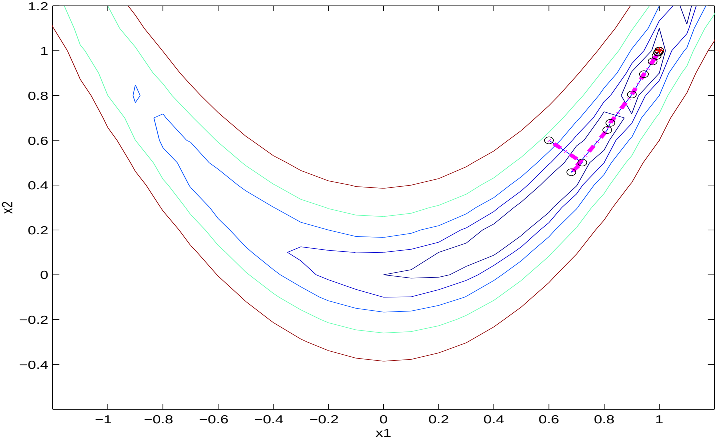
$$\min \; f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of Quasi-Newton algorithm (BFGS Method) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (-1.2, 1.0)^T$

Example (Rosenbrock function):

$$\min \quad f(\boldsymbol{x}) \stackrel{\Delta}{=} 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of Quasi-Newton algorithm (BFGS Method) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (0.6, 0.6)^T$

Broyden Family

- DFP method

$$\boldsymbol{B}_{DFP}^{k+1} = \boldsymbol{B}^k + \frac{\boldsymbol{\delta}^k \boldsymbol{\delta}^{k^T}}{\boldsymbol{\delta}^{k^T} \boldsymbol{\gamma}^k} - \frac{\boldsymbol{B}^k \boldsymbol{\gamma}^k \boldsymbol{\gamma}^{k^T} \boldsymbol{B}^k}{\boldsymbol{\gamma}^{k^T} \boldsymbol{B}^k \boldsymbol{\gamma}^k}$$

- BGFS method

$$\boldsymbol{B}_{BFGS}^{k+1} = \boldsymbol{B} + \left(1 + \frac{\boldsymbol{\gamma}^T \boldsymbol{B} \boldsymbol{\gamma}}{\boldsymbol{\delta}^T \boldsymbol{\gamma}}\right) \frac{\boldsymbol{\delta} \boldsymbol{\delta}^T}{\boldsymbol{\delta}^T \boldsymbol{\gamma}} - \left(\frac{\boldsymbol{\delta} \boldsymbol{\gamma}^T \boldsymbol{B} + \boldsymbol{B} \boldsymbol{\gamma} \boldsymbol{\delta}^T}{\boldsymbol{\delta}^T \boldsymbol{\gamma}}\right)$$

- *Broyden Family*

$$\boldsymbol{B}^{k+1}(\varphi) = \varphi \boldsymbol{B}_{BFGS}^{k+1} + (1 - \varphi) \boldsymbol{B}_{DFP}^{k+1}$$

where $\varphi \in [0, 1]$.

Quasi-Newton Algorithm (Broyden Family)

(1) Initialize $\boldsymbol{x}^0$, ϵ and symmetric positive definite $\boldsymbol{B}^0$, $\varphi \in [0, 1]$, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\boldsymbol{d}^k = -\boldsymbol{B}^k(\varphi)\boldsymbol{g}^k$

  (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that

    (i) $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) < f(\boldsymbol{x}^k)$

    (ii) α^k satisfies Armijo-Wolfe (or Armijo-Goldstein) conditions

  (c) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$

  (d) $\boldsymbol{B}^{k+1}(\varphi) = \varphi \boldsymbol{B}^{k+1}_{BFGS} + (1 - \varphi)\boldsymbol{B}^{k+1}_{DFP}$

  (e) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.