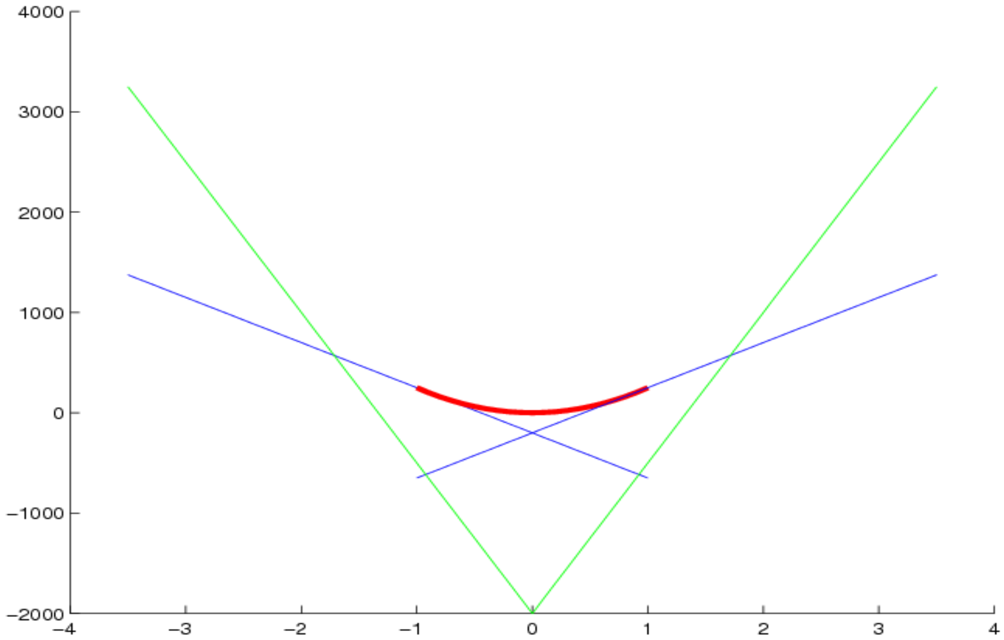
# Regularized Newton Method For Unconstrained Convex Optimization

Roman A. Polyak

George Mason University
USA

RPOLYAK@GMU.EDU



4000
3000
2000
1000
0
−1000
−2000
−4
−3
−2
−1
0
1
2
3
4

$f : R^n \to R$ Convex

$X^* = Ary\min\{f(x) | x \in R^n\} \neq \phi$ is bounded

$\|x\| = (x,x)^{\frac{1}{2}}$

$$F(x,y) = f(y) + 0.5\|\nabla f(x)\|\|y - x\|^2$$

1. $F(x,y)|_{y=x} = f(x)$

2. $\nabla_y F(x,y)|_{y=x} = \nabla f(x)$

3. $\nabla^2_{yy} F(x,y)|_{y=x} = \nabla^2 f(x) + \|\nabla f(x)\| I^n = H(x)$

if $x \notin X^*$ then there exists a unique

$$y(x) = \arg\min\{F(x,y) \mid y \in R^n\}$$

## Regularized Newton Step

$$x \in dom f(x) : 0 \le m(x) < M(x) < \infty$$

$$m(x)\|y\|^2 \le (\nabla^2 f(x)y, y) \le M(x)\|y\|^2$$

$$\hat{x} := x - (H(x))^{-1}\nabla f(x)$$

$$\hat{x} := x - (\nabla^2 f(x))^{-1}\nabla f(x)$$

$$r(x) = -(H(x))^{-1}\nabla f(x)$$

$$n(x) = -(\nabla^2 f(x))^{-1}\nabla f(x)$$

$$\nabla^2 f(x) = O^{n,n}$$

4. $$r(x) = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$$

$$f(y) - f(x) \geq (g(x), y - x)$$

$\nabla f(x) := g(x) -$ subgradient

5. $$r(x) = -\frac{g(x)}{\|g(x)\|}$$

$$d \in R^n$$

$$q(d) = -\frac{(\nabla f(x), d)}{\|\nabla f(x)\| \|d\|}$$

$$\max\{q(d) \mid \|d\| \le 1\} = q(-\frac{\nabla f(x)}{\|\nabla f(x)\|}) = 1$$

$$q(r(x)) = -\frac{(\nabla f(x), r(x))}{\|\nabla f(x)\| \|r(x)\|}$$

$$\nabla f(x) = -H(x) r(x)$$

$$q(r(x)) = \frac{(H(x) r(x), r(x))}{\|\nabla f(x)\| \|r(x)\|}$$

$$\geq \frac{(m(x) + \|\nabla f(x)\|) \|r(x)\|^2}{\|\nabla f(x)\| \|r(x)\|}$$

$$\|\nabla f(x)\| \leq \|H(x)\|\|r(x)\| \leq (M(x) + \|\nabla f(x)\|)\|r(x)\|$$

$$q(r(x)) \geq (m(x) + \|\nabla f(x)\|)(M(x) + \|\nabla f(x)\|)^{-1}$$

$$q(n(x)) \geq m(x)M^{-1}(x) = \kappa(x) \leq 1$$

$$\bar{q}(r(x)) - \bar{q}(n(x)) = \frac{m(x) + \|\nabla f(x)\|}{M(x) + \|\nabla f(x)\|} - \frac{m(x)}{M(x)}$$

$$= \frac{(1 - k(x))\|\nabla f(x)\|}{M(x) + \|\nabla f(x)\|}$$

## *Existence*

$n(x) \Leftrightarrow (\nabla^2 f(x))^{-1}$ for $\forall x \in X^*$

1. $r(x)$ always exists

2. $r(x)$ is always a descent direction

3. $\bar{q}(r(x)) - \bar{q}(n(x)) > 0, \qquad \forall x \in X^*$

## Regularized Newton Methods

$$\hat{x} = x - (\nabla^2 f(x) + \|\nabla f(x)\| I)^{-1} \nabla f(x)$$

$$S(x^*, \rho) = \{x : \|x - x^*\| \leq \rho\}$$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\| \quad (1)$$

$$m(y, y) \leq (\nabla^2 f(x) y, y) \leq M(y, y) \quad (2)$$

$$0 < m < M < \infty$$

$$r = (m + 0.5L)m^{-2}\|\nabla f(x)\| < 1$$

$$\rho \le m(m + 0.5L)^{-1} r$$

$$S(x^*, \rho) = \{x : \|x - x^*\|\} \le \rho$$

**Theorem** **If (1)-(2) are satisfied, then for any** $x^0 \in S(x^*, \rho)$

**the following bound holds**

$$\|x^s - x^*\| \le \frac{2m}{2m + L} r^{2^s}$$

$$\left\|\nabla f(\hat{x})\right\| \le \frac{m + 0.5L}{m^2}\left\|\nabla f(x)\right\|^2$$

$$= \frac{m^2}{m + 0.5L}\left(\frac{m + 0.5L}{m^2}\left\|\nabla f(x)\right\|\right)^2$$

$$r = \frac{m + 0.5L}{m^2}\left\|\nabla f(x)\right\| < 1$$

$$\left\|\nabla f(x^s)\right\| \le \frac{m^2}{m + 0.5L} r^{2^s}$$

$$\left\|\nabla f(x^{\circ})\right\| \geq m\left\|x^{\circ} - x^{*}\right\|$$

$$\left\|\nabla f(x^{\circ})\right\| = \frac{m^2 r}{m + 0.5L}$$

$$\left\|x^{\circ} - x^{*}\right\| \leq \frac{m}{m + 0.5L} r$$

$$\rho = \frac{m}{m + 0.5L} r$$

$$0 < \sigma < 0.5$$

$$\left\|\nabla f(\hat{x})\right\| \leq \frac{m + 0.5L}{m^2} \left\|\nabla f(x)\right\|^{\sigma} \left\|\nabla f(x)\right\|^{2-\sigma} \leq \left\|\nabla f(x)\right\|^{2-\sigma}$$

## Global Convergence of the Regularized Newton Method

$$\hat{x} := x - t(\nabla^2 f(x))^{-1} \nabla f(x) = x + tn(x)$$

$$f(x + tn(x)) \le f(x) + ct(\nabla f(x), n(x))$$

$(\nabla^2 f(x))^{-1}$ does not exists

$$\nabla^2 f(x) = 0^{n.n}$$

$$\hat{x} := x - t(\nabla^2 f(x) + \|\nabla f(x)\| I)^{-1} \nabla f(x)$$

$$\hat{x} := x - t\frac{\nabla f(x)}{\|\nabla f(x)\|} = x - t(x)\nabla f(x)$$

$$x^{s+1} = x^s - t(x^s)\nabla f(x^s)$$

**Regularized Newton** $\Rightarrow$ **gradient method**

$$\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\| \qquad (3)$$

$$\left(\nabla f(x) - \nabla f(y), x - y\right) \geq m\|x - y\|^2 \qquad m > 0 \qquad (4)$$

**B.T. Polyak 1963**

If (3) is satisfied, and $f(x) \geq f(x^*) > -\infty$ then for $0 < t(x_s) = t < 2M^{-1}$

(1) $f(x^s) > f(x^{s+1})$ and $\lim_{s \to \infty} \nabla f(x^s) = 0$

( 2 ) if (3)and (4) are satisfied then

$$0 < q(t) = 1 - 2tm + Mmt^2, \qquad q(M^{-1}) = \min_{t > 0} q(t) = 1 - mM^{-1}$$

$$\|x^s - x^*\|^2 \leq 2m^{-1}q^s(f(x^0) - f(x^*))$$

$\nabla f(x)$ does not exist

$\nabla^2 f(x) := 0^{n,n}$

$g(x) : f(y) - f(x) \geq \langle g(x), y - x \rangle$

always exists and for $x \notin X^*$ $g(x) \neq 0^n$

**Regularized Newton** $\Rightarrow$ **subgradient method**

$$\hat{x} := x - t \frac{g(x)}{\|g(x)\|}$$

**N. Z. Shor 1962, 1964**

$$x^{s+1} = x^s - t\, g(x^s)\left\|g(x^s)\right\|^{-1}$$

$$X_s = \left\{x : f(x) = f(x^s)\right\} \qquad \varepsilon > 0,\ \bar{s} > 0$$

$$\min_{x \in X_s} \left\|x - x^*\right\| \le 0.5t(1 + \varepsilon)$$

**Yu. M. Ermol'ev 1966, B.T. Polyak 1967**

$$\{t_s\} : a) t_s \to 0 \qquad b) \sum t_s = \infty$$

$$\varphi_s = \min_{1 \le i \le s} f(x^i) \qquad \lim_{s \to \infty} \varphi_s = f(x^*)$$

$$\lim_{s \to \infty} \min_{x \in X^*} \left\|x^s - x\right\| = 0$$

$$H(x)r(x) = -\nabla f(x) \qquad (5)$$

$$\nabla^2 f(x) = O^{n,n}$$

$$r(x) = \frac{\nabla f(x)}{\left\| \nabla f(x) \right\|}$$

If $\nabla f(x)$ does not exists then

$$r(x) = -\frac{g(x)}{\left\| g(x) \right\|}$$

$$\hat{x} := x + tr(x) \qquad (6)$$

## Algorithm (Global Regularized Newton Method)

### Initialization

Given accuracy $\epsilon > 0$ sequence $\{t_s\}_{s=0}^{\infty}$ which satisfies,

$$a) t_s \to 0$$

$$b) \sum t_s = \infty$$

small enough $\sigma > 0,\ m_0 > 0,$ large enough $M_0 > 0$

and a starting point $x^0 \in dom f(x)$

Set $\varphi := f(x^0), s := 0, \hat{\varphi} := \varphi, \hat{x} := x^0.$

step 1: $x := \hat{x}$, $\|\nabla f(x)\| \leq \epsilon$, then stop and output $x^* := x$.

step 2: If $\nabla^2 f(x)$ exists and $m(x) \geq m_0$, $M(x) \geq M_0$, then go to step 7

step 3: Set $\nabla^2 f(x) = O^{n,n}$ If $\nabla f(x)$ does not exist, then $\nabla f(x) := g(x)$.

Find $r(x)$ from (5).

step 4: Set $t := t_s, s := s+1$ and find $\hat{x}$ from (6).

step 5: If $f(\hat{x}) < \varphi$, then $\hat{\varphi} := f(\hat{x})$ go to step 1.

step 6: Set $x := \hat{x}$ go to step 3.

step 7: Find $r(x)$ from (5), set $t = 1$ and find $\hat{x}$ from (6)

step 8: If $\|\nabla f(\hat{x})\| > \|\nabla f(x)\|^{2-\sigma}$

Then $t := 0.5 m_0 M_0^{-1}$, find $\hat{x}$ from (6),

step 9: set $\hat{\varphi} := f(\hat{x})$ and go to step 1.

Theorem 2. If condition (1) and (2) are satisfied, then the grnm sequence converges to the unique solution $x^*$ from any starting point $x_0 \in domf(x)$ with asymptotic quadratic rate, i.e. for a given 0 < r < 1, there is $s_0 > 0$ such that

$$\left\| x^{s_0} - x^* \right\| \leq m(m + 0.5L)^{-1} r$$

and the following bound holds

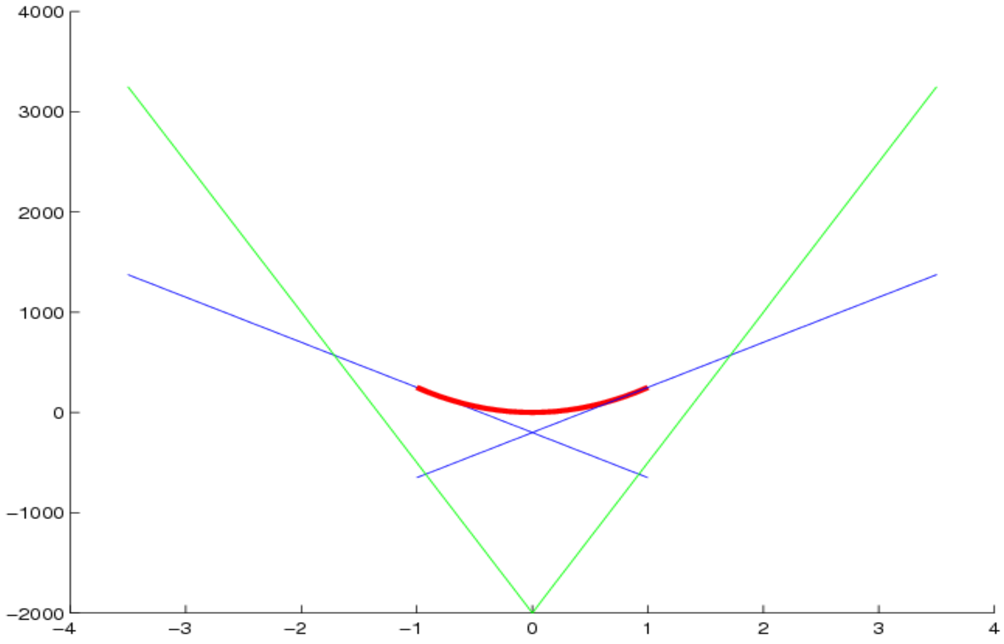
$$\left\| x^{s_0 + s} - x^* \right\| \leq \left(\frac{m}{m + 0.5L}\right)^2 r^{2^s + 1}$$

## Example

$$f_1(x) = \begin{cases} -3x-2, & -\infty \le x \le 1; \\ \frac{1}{2}(x^2+x^4), & -1 \le x \le 1; \\ 3x-2, & 1 \le x < \infty. \end{cases}$$

$$f_2(x) = \max\left\{\frac{16}{3}x-8, \frac{16}{3}x-8\right\}$$

$$f(x) = \max\{f_1(x), f_2(x)\}$$



4000
3000
2000
1000
0
−1000
−2000
−4
−3
−2
−1
0
1
2
3
4

**Example**

**Using the initialization values**:

$x^0 = 3.0,\ \sigma = 0.1,\ m_0 = 1,\ M_0 = 7$ and $t_s = s^{-1}$

The grmn generates the following sequence:

$$\left\{\begin{array}{l} 3.0;2.0;1.5;1.17;0.92;0.63; \\ 0.40;0.21;0.064;0.0063; \\ 6.6\times 10^{-5};1.2\times 10^{-8};8.3\times 10^{-16};\cdots \end{array}\right\}.$$