3. Trust-Region Methods

3.1

Rationale and idea: the subproblem.

- The problem with line-search methods. Very solid, (such as Armijo w. modified factorization), but one flaw. There is no way to distinguish a simple stationary point from SONC. It is unlikely, but...

- Alternative: Use second-order model in a region "of trust" and TAKE IT SERIOUSLY. that is, DO NOT MODIFY the matrix. If fail: reduce; if big success, enlarge.

- NOTATIONS:

  $f^k = f(x^k)$; $\nabla f^k = \nabla f(x^k)$;

3.1.2 Quadratic model:

$$m_k(p) = f^k + p^T g^k + \frac{1}{2} p^T B^k p$$

Taylor Series:

$$f(x^k + p) = f^k + p^T g^k + \frac{1}{2} p^T \nabla^2 f(x^k + tp) p$$

Difference: $m_k(p) - f(x^k + p) = O(\|p\|^2)$

[if $B^k = \nabla^2 f(x^k) \Rightarrow O(\|p\|^3)$]

Trust-Region Subproblem:

$$\boxed{\min_{p \in \mathbb{R}^n} m_k(p) \quad \|p\| \le \Delta^k}$$

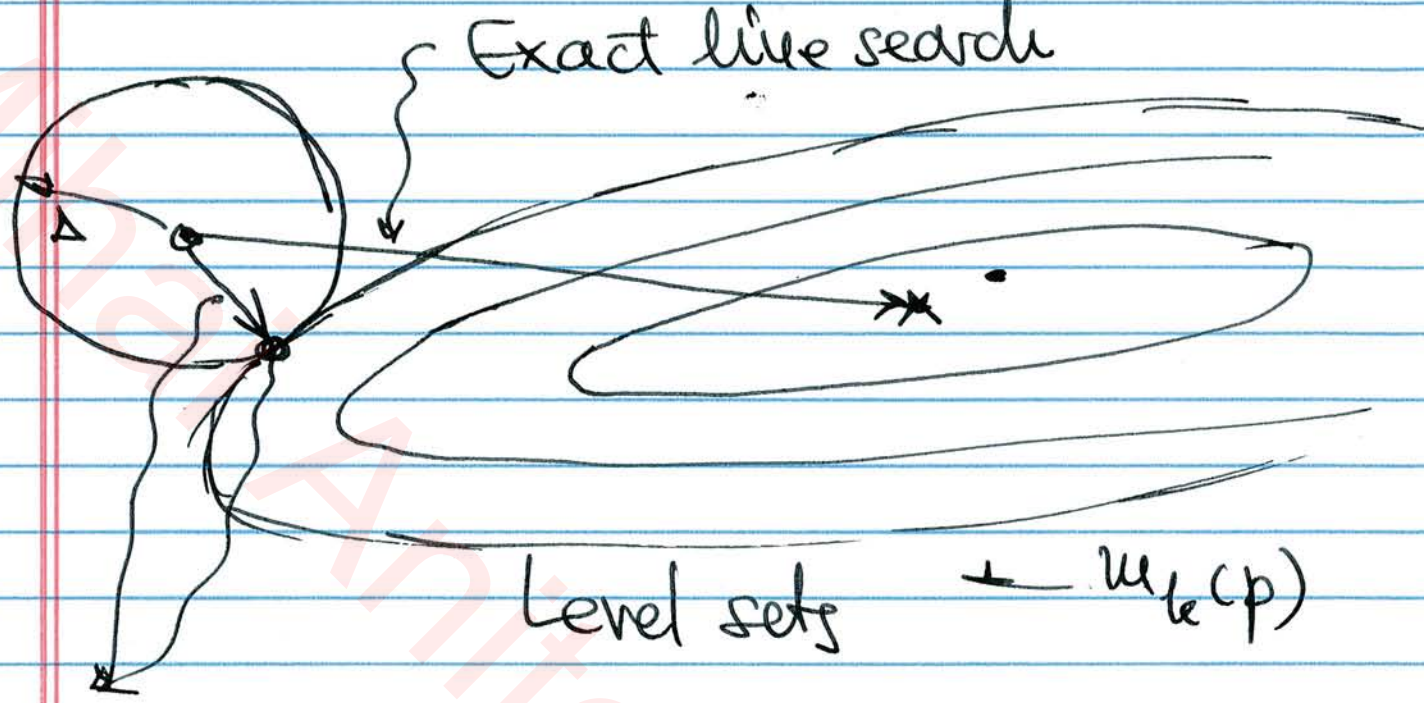
$\rightsquigarrow$ called the trust region

Discussion: If $B^k > 0$; and $\|(B^k)^{-1} g^k\| \le \Delta$

Then that is the solution. But the interesting case is elsewhere.

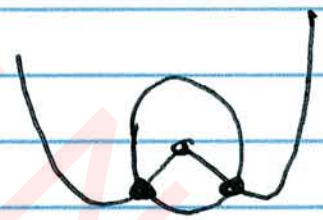
Mihai Anitescu Notes

3.1.3 ⊛ Intuition



Trust-region solution.

⊛ Note :- Example



$\cancel{((x^2-1)^2+1)}$

$(x^2-1)^2$ at 0. LS. is cooked

$2x(x^2-1)$

$4x^2-2$ ⇒ I escape saddle-point!

Mihai Anitescu Notes

3. Trust-Region Methods

3.2 Trust-Region Management

(*) How do we update TR size?

How well are we doing?

$$\rho^k = \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)}$$

← actual reduction

← predicted reduction

Both are > 0

Define $\eta \in [0, \frac{1}{4})$; $\hat{\Delta} > 0$

$\Delta_0 \in (0, \hat{\Delta})$

Obtain p^k by TR

Compute ρ^k

Δ^k manage

if $\rho^k < \frac{1}{4}$ → "too optimistic"

$$\Delta^{k+1} = \frac{1}{4}\Delta^k$$

else

if $\rho^k > 3/4$ and $\|p^k\| = \Delta^k$ → "too pessimistic"

$$\Delta^{k+1} = \min(2\Delta^k, \hat{\Delta})$$

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

Mihai Anitescu Notes

3.2.2

If $\rho^k \geq \eta$

$$x^{k+1} = x^k + p^k \quad \leftarrow \text{accept step}$$

else

$$x^{k+1} = x^k \quad \leftarrow \text{reject step}$$

How do we solve the TR subproblem

1) B^k is positive definite $\rightarrow$ LS, $\rightarrow$ NLE

"dogleg method"

2) B^k is not P.S.D., we have to mess around w. eigenvalues

Mihai Anitescu Notes

3. Trust-region

3.3. Cauchy Point

① Why: We need a convergence framework that allows us to play around with B^k.

⊛ Idea: $l_k(s) = f^k + g^{k^T} p$: linear model

Find: $$\boxed{p^{k,s} = \arg\min_{\substack{p \in \mathbb{R}^n \\ \|p\| \le \Delta^k}} f^k + g^{k^T} p}$$

Use $m_k(p)$ to minimize along $p^{k,s}$

$$\tau^k = \boxed{\arg\min_{\substack{\tau \ge 0 \\ \|\tau p^{k,s}\| \le \Delta^k}} m_k(\tau p_k^{k,s})}$$

Define $p^{k,c} = \tau^k p^{k,s}$

Note: Unless $g^k = 0$; the solution $p^{k,s} \neq 0$ and $\|p^{k,s}\| = \Delta^k$ $\Rightarrow$ $\boxed{p^{k,s} = \frac{-\Delta^k}{\|g^k\|} \cdot g^k}$

Mihai Anitescu Notes

3.3.2 Solution

$\alpha)$ $g^{k,T} B^k g^k \leq 0 \Rightarrow \tau^k = 1$

so $p^{k,s} = p^{k,c}$

$\beta)$ $g^{k,T} B^k g^k > 0 \Rightarrow$ quadratic equation

$$\tau^k = \begin{cases} 1 & g^{k,T} B^k g^k \leq 0 \\ \min\left(\|g^k\|^3 / (\Delta^k g^{k,T} B^k g^k), 1 \right) & \text{ow} \end{cases}$$

Idea: We converge if we use only Cauchy point!! Therefore the Cauchy point decrease becomes the yardstick by which we accept even incomplete TR solutions (cause they are hard).

Mihai Anitescu Notes

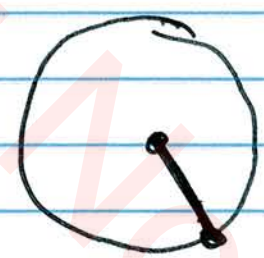
3. Trust-region approach; circumventing TR :-)

3.4 Dogleg Method

Q: How do we enhance upon the Cauchy point, and, hopefully, obtain superlinear convergence?

A: Go to Cauchy point, then move as far as you can on Newton direction. (makes sense really only for $B_k \succeq 0$ such as NE and LS)

The dogleg method.



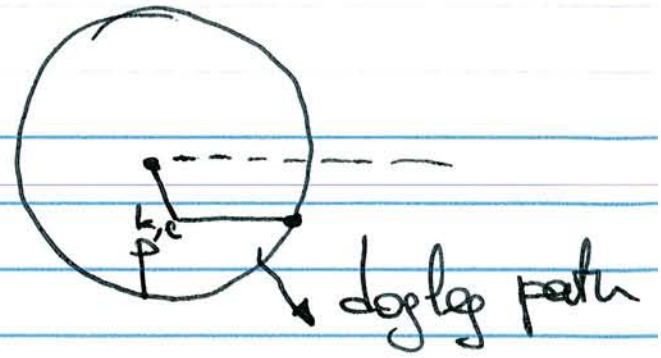
If [redacted] Cauchy point is interior (otherwise just stop at it)

$$\Rightarrow g^{k,T} B^k g^k > 0 \quad \text{and}$$

$$\boxed{\tau^k = \frac{\|g_k\|^3}{\Delta^k g^{k,T} B^k g^k}} \Rightarrow p^u = \tau^k p^{k,S} = \frac{\|g_k\|^{\cancel{3}}}{\cancel{\|\Delta^k\|} g^{k,T} B^k g^k} \cdot \frac{-\Delta^k g^k}{\cancel{\|g^k\|}}$$

Mihai Anitescu Notes

3.4.2 $$\boxed{p^u = -\frac{\|g^k\|^2}{g^{k,T} B^k g^k} \cdot g^k}$$



$$\boxed{p^B = -B_k^{-1} g^k}$$

dogleg path: $$\tilde{p}(\tau) = \begin{cases} \tau p^u & \tau \in (0,1) \\ p^u + (\tau - 1)(p^B - p^u) & 1 \le \tau \le 2 \end{cases}$$

If sufficiently close to solution, it will allow me to choose $\tau = 2$; the full Newton step and allow for SLC.

Lemma 4.2

Let B be positive definite. Then

i) $\|\tilde{p}(\tau)\|$ is an increasing function of τ

"moves away" from center

ii) $m_k(\tilde{p}(\tau))$ is a decreasing fcn. of τ

Mihai Anitescu Notes

3.4.3 <u>Proof</u>. For $\tau \in [0,1]$ the statement holds $\Rightarrow$ from Cauchy point properties

$$\boxed{\begin{aligned} h(\alpha) &= \frac{1}{2}\|\tilde{p}(1+\alpha)\|^2 = \\ &\frac{1}{2}\|p^u + \alpha(p^B - p^u)\|^2 = \frac{1}{2}\|p^u\|^2 + \\ &+ \alpha(p^u)^T(p^B - p^u) + \frac{1}{2}\alpha^2\|p^B - p^u\|^2 \end{aligned}}$$

$\Rightarrow$

$$h'(\alpha) = \alpha\|p^B - p^u\| + (p^u)^T(p^B - p^u)$$

$$\geq -(p^u)^T(p^u - p^B) = -\frac{(g^Tg)^3}{(g^TBg)^2} + \frac{g^TB^{-1}g \cdot g^Tg}{g^TBg}$$

$$= \frac{g^Tg}{(g^TBg)^2}\left[(g^TB^{-1}g)(g^TBg) - (g^Tg)^2\right] \geq 0$$

By Cauchy-Schwarz

Mihai Anitescu's notes

3.4.4 $\hat{h}(\alpha) = u(\tilde{p}(1+\alpha)) =$

$\cancel{f^k + \alpha g^{k,T} f}$

$$f + g^T(p^u + \alpha(p^B - p^u)) +$$

$$\frac{1}{2}(p^u + \alpha(p^B - p^u))^T B(p^u + \alpha(p^B - p^u))$$

$$\hat{h}' = g^T(p^B - p^u) + \cancel{\alpha(p^B - p^u)^T B(p^u + \alpha}$$

$$(p^B - p^u)^T B(p^u + \alpha(p^B - p^u)) \overset{\boxed{\alpha = 1}}{\leq}$$

$$(p^B - p^u)^T \cdot (\underbrace{g + Bp^B}_{=-g}) = 0.$$

QED

Mihai Anitescu Notes

3. TR methods

3.5 Convergence theory

⊛ Minimal decrease of Cauchy point

Lemma 4.3. $p^{k,c}$ satisfies

$$m_k(0) - m_k(p_k^c) \geq \frac{1}{2}\|g^k\| \min\left(\Delta^k, \frac{\|g^k\|}{\|B^k\|}\right)$$

<u>Proof</u>: Only one case: we prove:

$$g^T B g > 0 \quad \text{and} \quad \frac{\|g\|^3}{\Delta g^T B g} \leq 1 \quad (\zeta < 1)$$

$$\zeta = \frac{\|g\|^3}{\Delta g^T B g}$$

$$\Rightarrow \boxed{\cancel{m_k(p^c) - m_k(p^{k,c})}} \quad p^c = -\frac{\Delta g}{\|g\|}$$

$$m(p^c) - m(0) = g^T p^c + \frac{1}{2} p^{c,T} B p^c =$$

$$-\frac{\|g\|^4}{g^T B g} + \frac{1}{2}\cancel{\Delta^2} \cdot \frac{1}{\cancel{\|g\|^2}} \cancel{g^T B g} \cdot \frac{\|g\|^6}{\cancel{\Delta^2}(g^T B g)^{\cancel{2}}}$$

Mihai Anitescu Notes

3.5.B Theorem 4.5

Assume: 1. $\|B^k\| \leq \beta$; $\forall k$

DEFINE

2. $S = \{ x \mid f(x) \leq f(x^0) \}$

$$S(R_0) = \bigcup_{x \in S} \mathbb{B}(x \mid R_0)$$

f is B.B. on $S(R_0)$; for some $R_0 > 0$

3. ∇f is Lip on $S(R_0)$

4. We solve the TR subproblem $\boxed{\text{aproximately}}$ while satisfying:

$$(4.20): \; m_k(0) - m_k(p^k) \geq c_1 \cdot \|g^k\| \cdot \min\left(\Delta^k, \frac{\|g^k\|}{\|B^k\|}\right)$$

$$(4.25): \; \|p^k\| \leq \gamma \cdot \Delta^k \text{ for some } \gamma \leq 1$$

Then, $\liminf_{k \to \infty} \|g^k\| = 0$

Mihai Anitescu Notes

Intuition:

Assume:

1. $m_k(0) - m_k(p_k^S) \leq f(x^k) - f(x^{k+1})$
   + CP property
2. Δ_k is bounded below by $\bar{\Delta}$

$$\Rightarrow \sum \|g^k\| \cdot \min\left(\Delta^k, \frac{\|g^k\|}{\|B^k\|}\right) < \infty$$

$$\Rightarrow g^k \to 0$$

Mihai Anitescu Notes

3. Trust Region Methods

3.6. Iterative resolution of the subproblem

Theorem 4.1 : $\boxed{\text{Characterization of TR solution}}$

The vector p^* is a global solution of TR problem

$$\min_{p \in \mathbb{R}^n} m(p) = f + g^T p + \frac{1}{2} p^T B p \quad \text{st} \quad \|p\| \leq \Delta$$

$\Leftrightarrow$ p^* is feasible and $\exists\, \lambda \geq 0$ such that

$$(B + \lambda I)\, p^* = -g$$

$$\lambda(\Delta - \|p^*\|) = 0$$

$$(B + \lambda I) \text{ is P.S.D}$$

$\boxed{\text{Proof}}$

Mihai Anitescu Notes

3.6.1 Lemma 4.7. W.O Proof

Let $m(p) = g^T p + \frac{1}{2} p^T B p$; B symmetric

i) m attains minimum $\Leftrightarrow$ B is psd and g is in range of B

ii) m has a unique minimizer $\Leftrightarrow$ B is p.d.

<u>Proof, theorem</u>

a) $\exists\ \lambda \geq 0$ such that conditions are satisfied. $\overset{L.4.7}{\Rightarrow}$ p^* is a global min of

$$\hat{m}(p) = g^T p + \frac{1}{2} p^T (B + \lambda I) p =$$

$$m(p) + \frac{\lambda}{2} p^T p$$

Since $\hat{m}(p) \geq \hat{m}(p^*)$ $\Rightarrow$

Mihai Anitescu Notes

3.6.2

$$m(p) \geq m(p^*) + \frac{\lambda}{2}\left(p^{*T}p^* - p^T p\right)$$

$$\lambda \cdot \left(\|\Delta u - p^*\|\right) = 0 \quad \Rightarrow$$

$$m(p) \geq m(p^*) + \frac{\lambda}{2}\left(\Delta^2 - p^T p\right) \geq m(p^*)$$

↑ TR constraint

QED

$\Rightarrow$

Assume p^* is global solution

α) $\|p^*\| < \Delta$; verify yourselves by unconstrained minimization

$\lambda = 0$

β) $\|p^*\| = \Delta \quad \Leftrightarrow \quad p^T p = \Delta^2$

$\mathcal{L}(p, \lambda) = m(p) + \frac{\lambda}{2}\left(p^T p - \Delta^2\right)$ has a stationary point at p^*

Mihai Anitescu Notes

3.6.3 $\mathbb{B}p^* + g + \lambda p^* = 0 \Rightarrow (B + \lambda I)p^* = g$ (1)

(2) Since $m(p) \geq m(p^*)\ \forall p$ with $p^T p = \Delta^2$

$$m(p) \geq m(p^*) + \frac{\lambda}{2}\left((p^*)^T p^* - p^T p\right)$$

$$\cancel{g^T p + \frac{1}{2} p^T B p \geq g^T p^* +}$$

$$g^T p + \frac{1}{2} p^T B p \geq g^T p^* + \frac{1}{2} p^{*T} B p$$

$$+\lambda I \qquad\qquad +\lambda I$$

$$\boxed{g^T(p - p^*) =} -p^{*T}(B + \lambda I)(p - p^*) +$$

$$+ \frac{1}{2} p^T (B + \lambda I) p - \frac{1}{2} p^{*T}(B + \lambda I) p^* \geq 0 \quad \forall p$$

$$= \frac{1}{2} p^T (B + \lambda I) p + \frac{1}{2} p^{*T} (B + \lambda I) p^*$$

$$- \frac{1}{2} p^{*T} (B + \lambda I) p - \frac{1}{2} p^T (B + \lambda I) p^* =$$

Mihai Anitescu Notes

3.6.4 $\frac{1}{2}(p-p^*)^T(B+\lambda I)(p-p^*) \geq 0$

$\forall p; \|p\| \leq \Delta$

<u>Prove</u> Set of directions

$$\left\{ w : w = \pm \frac{p-p^*}{\|p-p^*\|} \text{ ; for some } p; \|p\| = \Delta \right\}$$

is dense on unit sphere $\Rightarrow$ P.S.D

(3) $\lambda \geq 0$

Because $\begin{cases} B+\lambda I \text{ is p.S.d.; and} \\ (B+\lambda I)p^* = -g \end{cases}$

$\Leftarrow$

<u>Lemma</u>

α) p^* is unconstrained min of

$$\hat{m}(p) = g^T p + \frac{1}{2} p^T (B+\lambda I) p$$

β) Point p^* in unc. min of $m(p)$

Mihai Anitescu Notes

3.6J From a):

$\hat{m}(p) \geq \hat{m}(p^*) \Rightarrow$ $\boxed{\forall p}$

$m(p) \geq m(p^*) + \frac{1}{2}\lambda(p^{*T}p^* - p^T p)$

Assume $\lambda \overset{\text{only}}{<} 0$ satisfies it.

$\Rightarrow \forall p;\ \|p\| \geq \Delta \Rightarrow$

$m(p) \geq m(p^*) + \frac{1}{2}\lambda(p^{*T}p^* - p^T p) \geq m(p^*)$

$\Rightarrow p^*$ is a global min. $\Rightarrow$

$Bp = -g$; B is p.s.d $\Rightarrow$ $\lambda = 0$ works

(Lemma).

$\boxed{\Sigma}$

Mihai Anitescu Notes

3.6.6 Solving the TR problem.

1) $\|p\| \le \Delta$; $\lambda = 0$; or

$$\boxed{2)\ p(\lambda) = -(B + \lambda I)^{-1} g}$$

$B + \lambda I$ is P.D.

& $\|p(\lambda)\| = \Delta$.

Since B is symmetric ; $\exists$ Q orthogonal

$Q = [q^1, q^2, \dots q^n]$; such that

$$B q^i = \lambda^i q^i \Rightarrow BQ = Q\Lambda$$

$$\lambda^1 \le \lambda^2 \le \dots \lambda^n$$

$$\Rightarrow B + \lambda I = Q(\Lambda + \lambda I)Q^T$$

$$p(\lambda) = -(B + \lambda I)^{-1} \cdot g =$$

Mihai Anitescu Notes

3.6.7

$$= -Q(\Lambda+\lambda I)Q^T g =$$

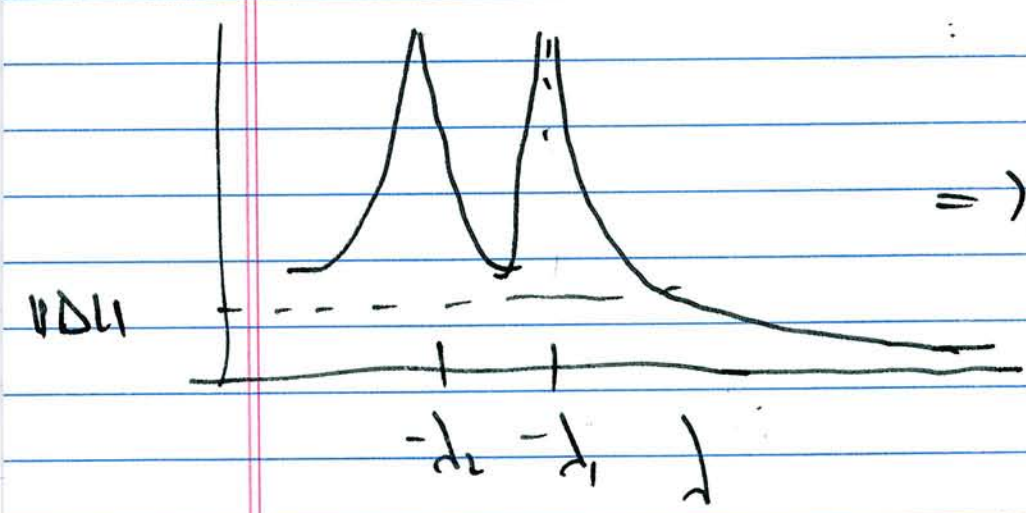
$$= -\sum_{j=1}^{n} \frac{q^{jT} g}{(\lambda+\lambda^j)} q^j$$

$$\Rightarrow \|p(\lambda)\|^2 = \sum_{j=1}^{n} \frac{(q^{jT} g)^2}{(\lambda+\lambda^j)^2} = \Phi(\lambda)$$

Observations

1) $\lambda > -\lambda^1$; $\Phi(\lambda)$ is decreasing

2) $q^{1T} g \neq 0 \Rightarrow \lim_{\lambda \to \rho} \|p(\lambda)\| = +\rho$



$\Rightarrow$ $\boxed{\text{MUST CROSS } \Delta.}$

Mihai Anitescu Notes

3.6.8 $\boxed{\text{Find } \phi_1(\lambda)=0 \quad \lambda \in [-\lambda_1, \infty).}$

For stability Purposes; Solve as

$$\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}$$ (Note; diff!! at $\lambda > -\lambda_1$)

Use newton

$$\lambda^{\ell+1} = \lambda^{\ell} - \frac{\phi_2(\lambda^{(\ell)})}{\phi_2'(\lambda^{\ell})}$$

$\Rightarrow$ $\boxed{\text{Algorithm 4.3 (using Cholesky)}}$

Factor $B + \lambda^{(\ell)} I = R^T R$

Solve $R^T R\, p^{\ell} = -g$; $R^T q^{\ell} = p^{\ell}$

Set $\lambda^{(\ell+1)} = \lambda^{(\ell)} + \left(\frac{\|p^{\ell}\|}{\|q^{\ell}\|}\right)^2 \left(\frac{\|p^{\ell}\| - \Delta}{\Delta}\right)$.

Mihai Anitescu Notes

3.6.9 Problems: Need safeguards. Need estimate of λ_1. See More-Sorensen

Hard Case

$q_1^T g = 0 \quad ; \quad \lambda_1 < \lambda_2 \leq \ldots$

$$\Rightarrow p(\lambda) = -\sum \frac{q^{j^T} g}{\lambda + \lambda^j} q^j$$

$\Rightarrow$ ~~We cannot~~ What if $\|p(-\lambda_1)\| < \|\Delta\|$

$\Rightarrow B + \lambda I$ is invertible for $\lambda \in (-\lambda_1, \infty)$

$\Rightarrow \|p(\lambda)\| < \|\Delta\| \Rightarrow$ no solution

$\Rightarrow$ $\boxed{\text{WE MUST HAVE } \lambda = -\lambda_1}$

$(B + \lambda_1) p = -g \quad \Rightarrow \quad p = p + \tau q^1$

is also a solution

Mihai Anitescu Notes

3.6.10. $\hat{p}(\lambda) = p(\lambda_1) + z q^1 \Rightarrow$ Choose z

such that $\|\hat{p}(\lambda)\| = 0$

(TEST $(B + \lambda_1 I) \cdot \hat{p}(\lambda) = q^1$.)

Because $p(\lambda_1) \perp q^1 \Rightarrow$

$\|\hat{p}(\lambda)\|^2 = \|p(\lambda_1)\|^2 + z^2 \Rightarrow$ Two Choices

$\boxed{\text{Extend to multiple eigs...}}$

Mihai Anitescu Notes

3. Trust-region methods

3.8. Convergence of algorithms

Observation: Need not solve TR exactly !!

Approximate resolution.

$$m(0) - m(p) \geq c_1 [m(0) - m^*(p^*)]$$

$$\|p\| \leq \gamma \Delta. \qquad \gamma > 1$$

Theorem 4.8.

Assume that: 1) $\eta \in (0, 1/4)$

2) $\|B_k\| \leq \beta$ for some β

3) f is bdd below on S

4) f is Lipsch cts diff on $S(R_0)$

5) AR holds

6) $B_k = \nabla^2 f(x^k)$

Then

$$\boxed{\lim_{k \to \infty} \|g^k\| = 0}$$

Mihai Anitescu Notes

7)

3.8.2 If, in addition, level set is compact

then either α) x^k ~~$\to x^*$~~ terminates at x^*, SONC

β) $\{x^k\}$ has a limit point x^* SONC.

~~Theorem 4.9~~ :

Theorem 4.9 : "In addition, superlinearly convergent".

{ Idea: If Δ^k not too small; $B^k > 0$

$\Rightarrow$ choose Newton step

Mihai Anitescu Notes

3. Trust Region

3.9 Enhancements

Scaling:

$$\min_{p \in \mathbb{R}^n} \quad m_k(p) = f^k + g^{k^T} p + \frac{1}{2} p^T B^k p$$

$$\text{s.t.} \quad \|Dp\| \le \Delta^k$$

$$\Leftrightarrow \quad Dp = r$$

$$\boxed{\begin{aligned} \min_{r \in \mathbb{R}^n} \quad & m_k(r) = f^k + g^{k^T} D^{-1} r + \\ & + \frac{1}{2} r^T D^{-1} B^k D^{-1} r \end{aligned}}$$

$$\|r\| \le \Delta^k$$

a good choice:

$$\boxed{D^{-1} B D^{-1} \sim I}$$

Mihai Anitescu Notes

3. TRUST REGION

3.10 LS v. TR.

	LS	TR
SoNC	Cannot hope to achieve SoNC $\nabla f(x^*) = 0$	Yes, sometimes $\nabla f(x^*) = 0$

	LS	TR
SoNC	N	Y, under some assumption
FoNC	Y (Z)	Y : Thm. 4.5, 4.6, 8
SLC	Y if Newton	Y if Newton
SUBPROBLEM	LE, but for which matrix.	TR, iterative, harder

Milan Antiesch Notes