Efficient Solutions for the Large Scale Trust Region Subproblem, TRS

Henry Wolkowicz

(work with Ting Kei Pong and Heng Ye)
Dept. of Combinatorics and Optimization
University of Waterloo

WCOM FALL, 2011, at UBC Okanagan, Kelowna, BC

Outline

Background

- Generalized Trust Region Subproblem and Characterizations of Optimality
- MoSo (1983) algorithm for TRS; within a primal-dual framework; handling the hard case

Duality/Exploiting Sparsity & Degeneracy/Hard Case

- Various primal-dual pairs with strong duality
- Parametric eigenvalue problem and sparsity
- Hard case is actually easiest case, with an explicit solution
- Underdetermined least squares, Numerics

Trust Region Subproblem, (TRS)

“Nonconvex” quadratic minimization

$$\text{(TRS)} \qquad \begin{array}{rll} q^* = & \min & q(x) := x^T A x - 2a^T x \\ & \text{s.t.} & \|x\| \leq s, \\ & & x \in \mathbb{R}^n \end{array}$$

- $A \in \mathcal{S}^n$ - $n \times n$ symmetric (possibly indefinite) matrix
- $a \in \mathbb{R}^n$; $s > 0$ (TR radius);

Generalized TRS, (GTRS)

$$\text{(GTRS)} \qquad \begin{array}{rll} q^* = & \min & q(x) := x^T A x - 2a^T x \\ & \text{s.t.} & \ell \leq q_1(x) := x^T B x - 2b^T x \leq u \end{array}$$

Applications

Many important applications:

- forming subproblems for constrained optimization
- regularization of ill-posed problems
- theoretical applications
- **trust region (TR) methods for unconstr. min.** uses TRS formed from second order Taylor series approximation

(dis)Advantages for TR approach

- obtains second order optimality conditions
- q-quadratic convergence
- **BUT:** popularity?:
  (i) difficulties exploiting sparsity?
  (ii) hard case (degenerate case)?
  (iii) evaluating Hessian?

Characterization of Optimality for TRS

Surprising; TRS is on *watershed* of convex/nonconvex

- a characterization of optimality of a (possibly) nonconvex problem
- second order positive semidefinite necessary conditions hold on all of $\mathbb{R}^n$

Characterization of x^* optimal for TRS (iff)

(Gay-81 [2], More-Sorensen-83 [6])

$$\left.\begin{array}{c} (A-\lambda^* I)x^* = a, \\ A-\lambda^* I \succeq 0, \lambda^* \leq 0 \end{array}\right\} \qquad \text{dual feasibility}$$

$$\|x^*\|^2 \leq s^2 \qquad \text{primal feasibility}$$

$$\lambda^*(s^2 - \|x^*\|) = 0 \qquad \text{complementary slackness}$$

Characterization of Optimality for (homog) GTRS

(Mild) Slater Type Constraint Qualification

$$\text{(GTRS)} \qquad \begin{array}{rll} q^* = & \min & q(x) := x^T A x - 2a^T x \\ & \text{s.t.} & \ell \leq q_1(x) := x^T B x \leq u \end{array}$$

$$(CQ) \qquad \exists \hat{x} \text{ s.t. } l < q_1(\hat{x}) := \hat{x}^T B \hat{x} < u$$

Characterization of x^* optimal for GTRS (iff)

(More-93 [5], Stern-W-93 [10])

$$\left.\begin{array}{c} (A - \lambda^* B) x^* = a, \\ A - \lambda^* B \succeq 0 \end{array}\right\} \qquad \text{dual feasibility}$$

$$\ell \leq {x^*}^T B x \leq u \qquad \text{primal feasibility}$$

$$\lambda^*(u - {x^*}^T B x^*) \geq 0 \geq \lambda^*({x^*}^T B x^* - \ell) \qquad \text{compl. slack.}$$

S-Lemma Yakubovich-71,73

(survey in Polik-Terlaky:07 [7])

$f, g : \mathbb{R}^n \to \mathbb{R}$ quadratic functions
$\exists \bar{x}$ s.t. $g(\bar{x}) < 0$
then TFAE:

1. $\nexists x \in \mathbb{R}^n : f(x) < 0, g(x) \leq 0$
2. $\exists \lambda \geq 0$ s.t. $f(x) + \lambda g(x) \geq 0, \forall x \in \mathbb{R}^n$

Finsler:37 , Hestenes-McShane:40
$\{x^T Bx = 0, x \neq 0 \implies x^T Ax > 0\}$
implies
$\{\exists \lambda \geq 0, A + \lambda B \succ 0\}$

Mo-So Algorithm Framework

Recall: Opt. conditions of TRS

$$\left.\begin{array}{c}(A-\lambda^* I)x^* = a, \\ A-\lambda^* I \succeq 0, \lambda^* \leq 0\end{array}\right\} \qquad \text{dual feasibility}$$

$$\|x^*\|^2 \leq s^2 \qquad \text{primal feasibility}$$

$$\lambda^*(s^2 - \|x^*\|) = 0 \qquad \text{complementary slackness}$$

Original view: if $A - \lambda^* I \succ 0, \lambda^* < 0$

- define: $x(\lambda) = (A - \lambda I)^{-1} a$
- <u>SOLVE</u>: $\psi(\lambda) := \|x(\lambda)\| - s = \|(A - \lambda I)^{-1} a\| - s = 0$
- maintain: $A - \lambda I \succ 0, \ \lambda \leq 0$

Less Nonlinear $\phi(\lambda)$ for Newton's Method

Newton for $\psi(\lambda) = 0$?

$A = Q\Lambda Q^T$ orthogonal diagonalization; $\gamma = Q^T a$;
$\psi(\lambda) = \|x(\lambda)\|^2 - s^2 = \sum_{j=1}^{n} \frac{\gamma_j^2}{(\lambda_j(A)-\lambda)^2} - s^2$ *highly* nonlinear in λ, in particular near $\lambda_1(A)$.

less nonlinear $\phi(\lambda)$ (Reinsch:67 [8, 9], Hebden:73 [3])

SOLVE: $\phi(\lambda) := \frac{1}{s} - \frac{1}{\|x(\lambda)\|} = 0$

Newton iterates $\lambda^{(k)}$ for $\phi(\lambda) = 0$; exploit Cholesky

$$\begin{aligned}
\lambda^{(k+1)} &= \lambda^{(k)} - \frac{\phi(\lambda^{(k)})}{\phi'(\lambda^{(k)})} \\
&= \lambda^{(k)} - \left(\frac{\|x\|}{\|y\|}\right)^2 \left(\frac{\|x\|-s}{s}\right),
\end{aligned}$$

where: $A - \lambda^{(k)} I = R^T R$ (<u>Cholesky</u>); $R^T R x = a$; $R^T y = x$

Computing the Newton direction/step Gay:81, MoSo:83

Assume $\lambda^{(k)} \leq 0$ and $A - \lambda^{(k)} I \succ 0$ (i.e. $\lambda^{(k)} < \lambda_1(A)$)

- Factor $A - \lambda^{(k)} I = R^T R$ (Cholesky factorization).
- Solve for x, $R^T R x = a$ $\quad$ ($x = x(\lambda^{(k)})$).
- Solve for y, $R^T y = x$.
- Let $\lambda^{(k+1)} = \lambda^{(k)} - \left[\frac{\|x\|}{\|y\|}\right]^2 \left[\frac{(\|x\| - s)}{s}\right]$ $\quad$ (Newton step); but, *safeguard/maintain* positive definiteness

Nearly linear

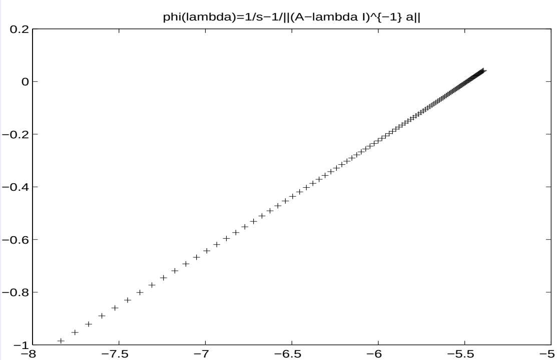


Figure: “Easy” case: $\phi(\lambda) = \frac{1}{s} - \frac{1}{||(A-\lambda I)^{\dagger} a||}$

“Easy” Case; Newton Quadr. Cvgnce from One Side

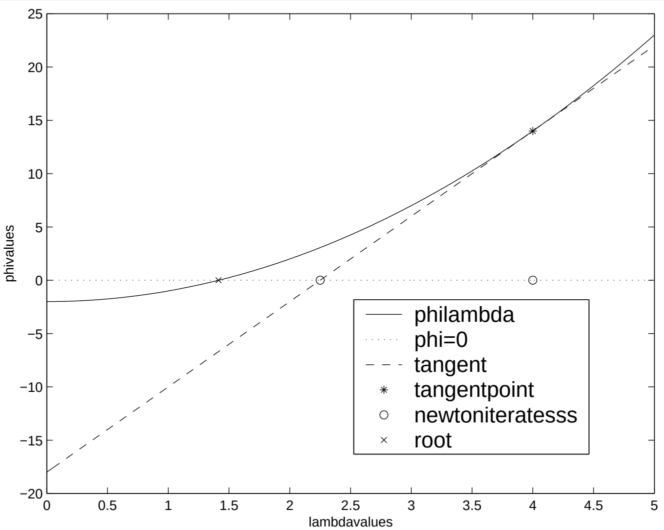


Figure: Newton’s method with the secular function, $\phi(\lambda)$.

Easy/Hard Cases for TRS

$\lambda_1(A) = \lambda_{\min}(A) > \lambda^* \implies A - \lambda^* I \succ 0$

1. Easy case	2.(a) Hard case (case 1)	2.(b) Hard case (case 2)
$a \notin \mathcal{R}(A - \lambda_1(A)I)$ (stationarity implies $\lambda^* \neq \lambda_1(A)$, so that $\lambda^* < \lambda_1(A)$)	$a \perp \mathcal{N}(A - \lambda_1(A)I)$ but $\lambda^* < \lambda_1(A)$	$a \perp \mathcal{N}(A - \lambda_1(A)I)$ and $\lambda^* = \lambda_1(A)$ (i) $\|(A - \lambda^* I)^\dagger a\| = s$ or $\lambda^* = 0$ (ii) $\|(A - \lambda^* I)^\dagger a\| < s, \lambda^* < 0$

Table: The three different cases for the trust region subproblem. We include two subcases (i) and (ii) for the hard case (case 2), where $A - \lambda^* I$ is positive semidefinite, singular. ($\cdot^\dagger$ denotes Moore-Penrose gen. inverse.)

Hard Case is really easiest?

We take advantage of Lanczos algorithm used in large sparse case.
We will see: hard case becomes easiest case.

$\phi(\lambda)$ in the (near) Hard Case

Newton steps (linearizations) are *"useless"*

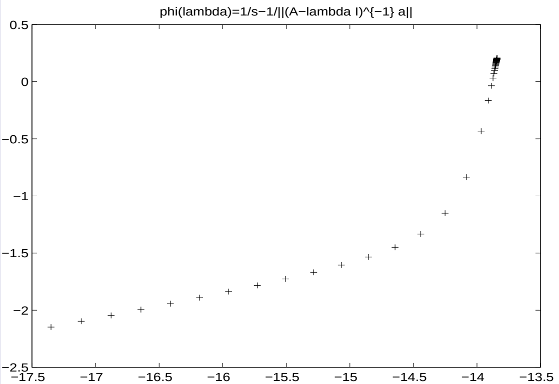


Figure: (near) hard case case: $\phi(\lambda) = \frac{1}{s} - \frac{1}{||(A-\lambda I)^{\dagger} a||}$

Nonlinear Primal-Dual Pair of SDPs; MoSo Alg.

Lagrangian: $L(x,\lambda) := x^tAx - 2a^tx - \lambda(||x||^2 - s^2)$

(Stern-W.:93 [10])

- Strong duality holds for (TRS)
  $$\mu^* = \min_x \max_\lambda L(x,\lambda) = \max_\lambda \min_x L(x,\lambda).$$
  Moreover, attainment holds for x and uniquely for λ.
- A (Wolfe) dual problem for (TRS), without a duality gap:
  $$(D) \qquad \mu^* = \max_{A-\lambda I \succeq 0} h(\lambda), \quad \text{where}$$
  $$h(\lambda) = \lambda s^2 - a^t(A-\lambda I)^\dagger a$$

MoSo is a dual algorithm

solves a modified:
$$h'(\lambda) = s^2 - a^t\left((A-\lambda I)^\dagger\right)^2 a = s^2 - \|x(\lambda)\|^2 = 0$$
a duality gap is: $q(x_{feas}) - h(\lambda_{feas})$

Easy case: $h(\lambda) = \lambda s^2 - a^t(A - \lambda I)^\dagger a$

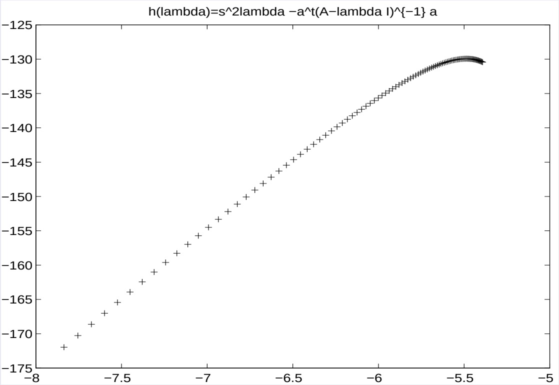


Figure: Easy case: $h(\lambda) = \lambda s^2 - a^t(A - \lambda I)^\dagger a$

Easy case, solve: $h'(\lambda) = s^2 - a^t((A - \lambda I)^\dagger)^2 a = 0$

Can be very nonlinear

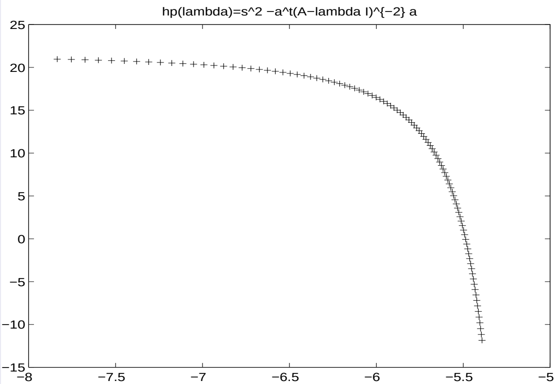


Figure: Easy case: $h'(\lambda) = s^2 - a^t((A - \lambda I)^\dagger)^2 a = s^2 - \|x(\lambda)\|^2$

(near) Hard case: $h(\lambda) = \lambda s^2 - a^t(A - \lambda I)^\dagger a$

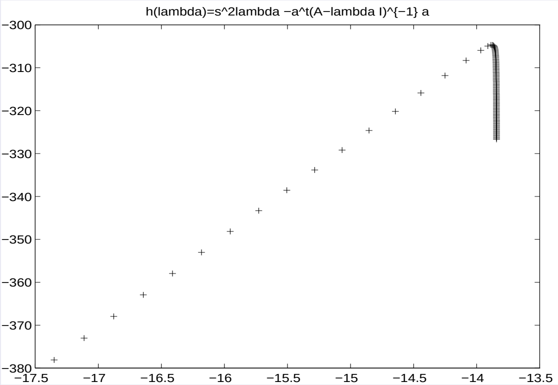


Figure: (near) Hard case: $h(\lambda) = \lambda s^2 - a^t(A - \lambda I)^\dagger a$

Dual to SDP

Slater's CQ holds for dual program (D)

Strong duality holds with the (Wolfe) dual (DD) of the dual (D):

$$
(DD) \qquad \begin{array}{rcl} \mu^* = & \min & h(\lambda) + \mathrm{trace}\, X(A-\lambda I) \\ & \text{s.t.} & s^2 - \|(A-\lambda I)^\dagger a\|^2 - \mathrm{trace}\, X = 0 \\ & & X \succeq 0. \end{array}
$$

Duality gap view to handling hard case in MoSo

$h(\lambda) + \mathrm{trace}\, X(A-\lambda I) - h(\lambda) = \mathrm{trace}\, X(A-\lambda I)$ (compl. slack.)

if $\|x_\lambda\| < s$ and $\|x_\lambda + z\|^2 = s^2$, then:

$$
\begin{array}{rcl} q(x_\lambda + z) & = & (\lambda s^2 - x_\lambda^t (A-\lambda I) x_\lambda) + z^t (A-\lambda I) z \\ & = & h(\lambda) + z^t (A-\lambda I) z. \end{array}
$$

find z (a direction of negative curvature) to make $z^t(A-\lambda I)z$ small. Set $X = zz^t$ to make duality gap term, or complementary slackness term, small. (Error estimate now comes from duality gap.)

Towards a Parametrization for Large Sparse Case

Motivation: try to avoid Cholesky in large sparse case

Homogenization, $y_0^2 = 1 \quad \mu^* =$

$$
\begin{array}{rl}
= & \min\limits_{||x||=s,\ y_0^2=1} x^tAx - 2y_0a^tx \\
= & \max\limits_{t} \min\limits_{||x||=s,\ y_0^2=1} x^tAx - 2y_0a^tx + ty_0^2 - t \\
\geq & \max\limits_{t} \min\limits_{||x||^2+y_0^2=s^2+1} x^tAx - 2y_0a^tx + ty_0^2 - t \quad \text{**eig prob**} \\
\geq & \max\limits_{t,\lambda} \min_{x,y_0} x^tAx - 2y_0a^tx + ty_0^2 - t + \lambda(||x||^2 + y_0^2 - s^2 - 1) \\
= & \max\limits_{r,\lambda} \min_{x,y_0} x^tAx - 2y_0a^tx + ry_0^2 - r + \lambda(||x||^2 - s^2) \\
= & \max\limits_{\lambda} \left(\max\limits_{r} \min_{x,y_0} x^tAx - 2y_0a^tx + ry_0^2 - r + \lambda(||x||^2 - s^2) \right) \\
= & \max\limits_{\lambda} \min_{x,y_0^2=1} x^tAx - 2y_0a^tx + \lambda(||x||^2 - s^2) \\
= & \mu^*,
\end{array}
$$

equated r with $t + \lambda$.

used strong duality for TRS for last two equalities.

Parametric Eigenvalue Problem

$$\max_t \boxed{\min_{||x||^2+y_0^2=s^2+1} x^tAx - 2y_0a^tx + ty_0^2} - t$$

$$k(t) := (s^2+1)\lambda_1(D(t)) - t \qquad D(t) = \begin{bmatrix} t & -a^t \\ -a & A \end{bmatrix}$$

An Unconstrained dual problem to (TRS)

$$\max_t k(t) \qquad \text{(concave max)}$$

$k'(t) = (s^2+1)y_0(t)^2 - 1$

If $\lambda_1(D(t))$ is simple

- $y(t)$ normalized eigenvector for $\lambda_1(D(t))$
- $y(t) = \begin{pmatrix} y_0(t) \\ x(t) \end{pmatrix}$
- $$\frac{1}{y_0(t)}\|x(t)\| = s \text{ if and only if } k'(t) = 0$$

Linear Primal-Dual SDP Pair

From parametric eigenvalue problem

$$(DSDP) \qquad \begin{array}{rll} \mu^* = & \max & (s^2+1)\lambda - t \\ & \text{s.t.} & D(t) \succeq \lambda I \end{array}$$

(with $\lambda \leq 0$ for inequality constraint TRS)

$$(PSDP) \qquad \begin{array}{rll} \mu^* = & \min & \operatorname{trace} D(0)X \\ & \text{s.t.} & \operatorname{trace} X = s^2+1 \\ & & X_{11} = 1 \\ & & X \succeq 0 \end{array}$$

$y_0(t)$ in the easy case

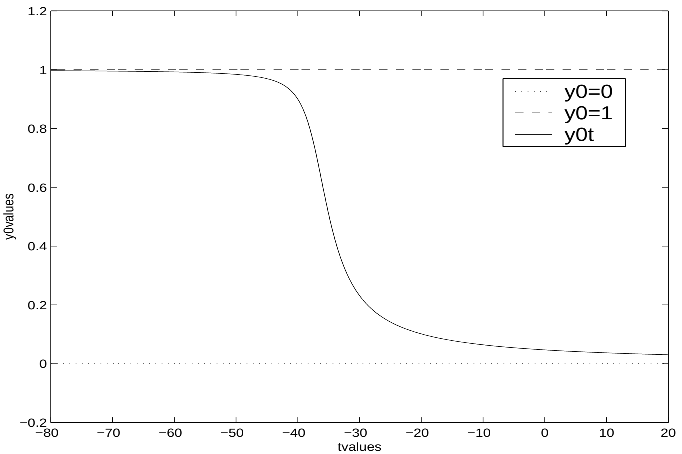


Figure: $y_0(t)$ in the easy case

$y_0(t)$ in the hard case

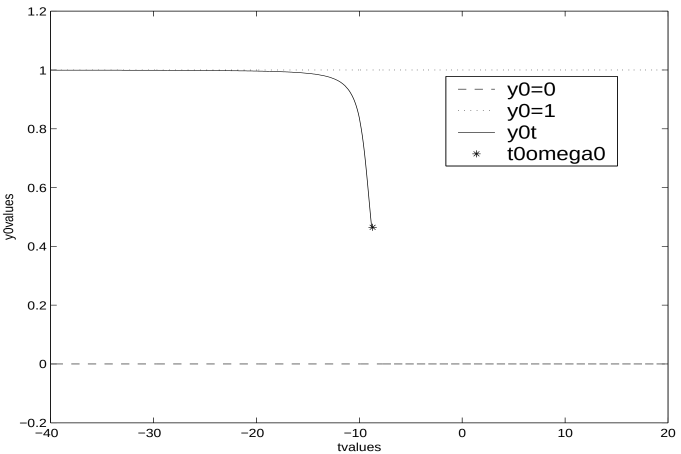


Figure: $y_0(t)$ in the hard case

$k'(t)$ in the easy case

Can be highly nonlinear

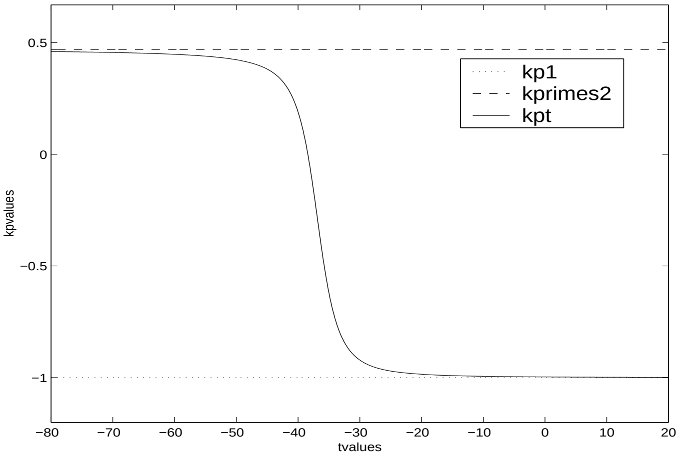


Figure: $k'(t)$ in the easy case

$k'(t)$ in the hard case (*case 1*)

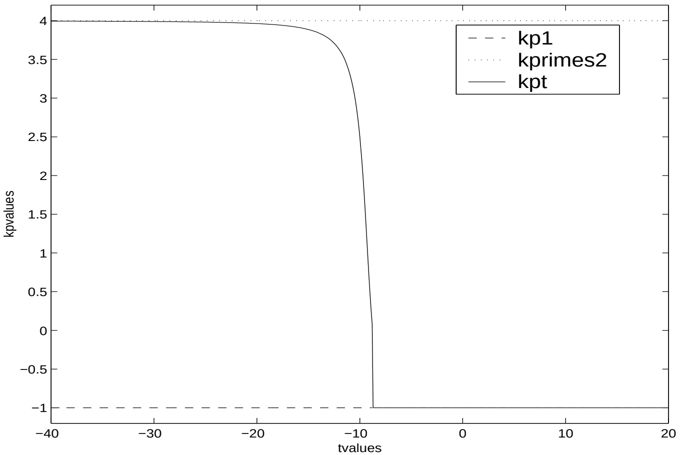


Figure: $k'(t)$ in the hard case (*case 1*)

Solving $\psi(t) = 0$ or $k'(t) = 0$ is equivalent; but less nonlinear.

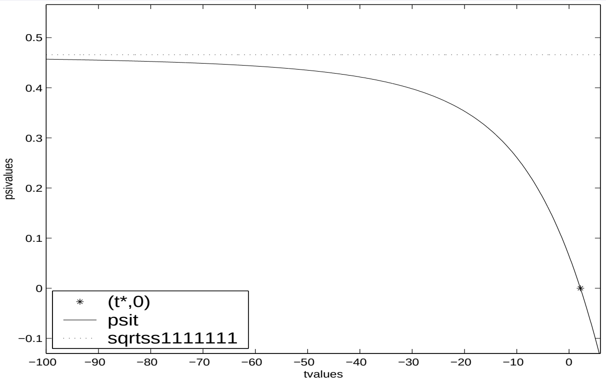


Figure: $\psi(t)$ in the easy case

$\psi(t)$ in the hard case (*case 1*)

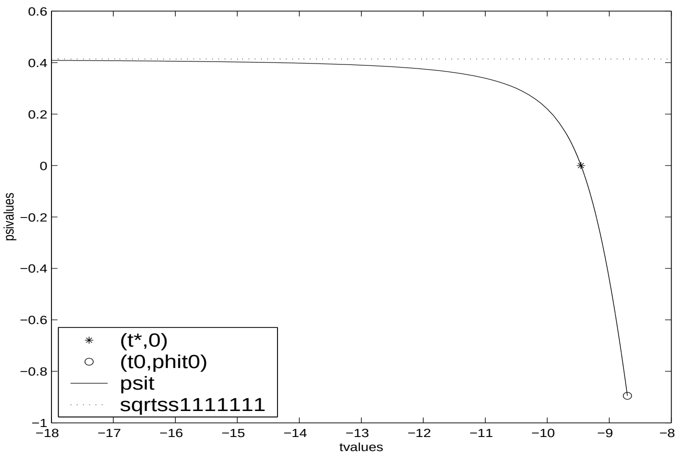


Figure: $\psi(t)$ in the hard case (*case 1*)

$\psi(t)$ in the hard case (*case 2*)

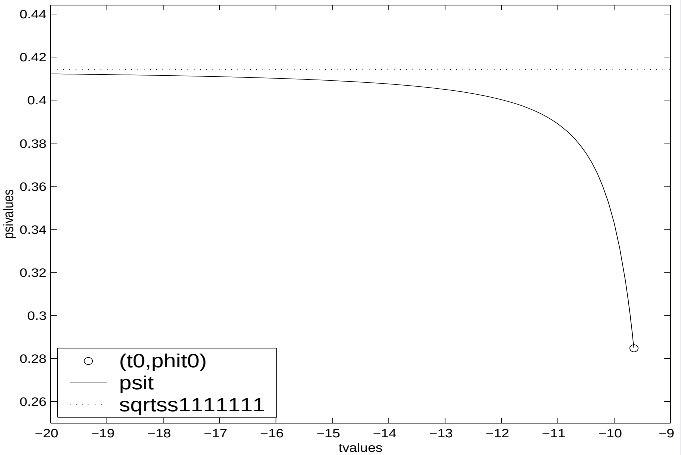


Figure: $\psi(t)$ in the hard case (*case 2*)

easy case: $k(t) = (s^2 + 1)\lambda_1(D(t)) - t$

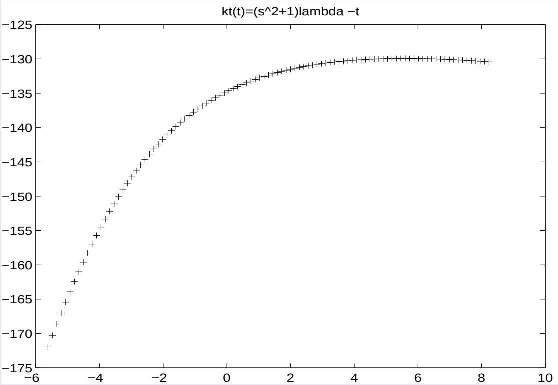


Figure: easy case: $k(t) = (s^2 + 1)\lambda_1(D(t)) - t$

$k(t)$ Hard Case (Case 1)

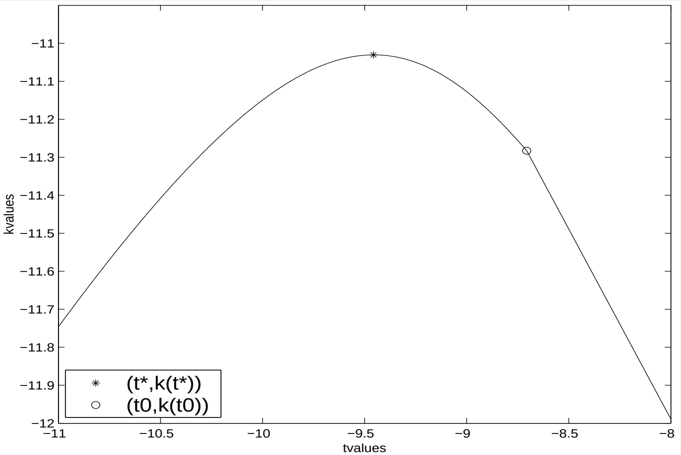


Figure: $k(t)$ in the hard case (case 1)

$k(t)$ in the hard case (case 2)

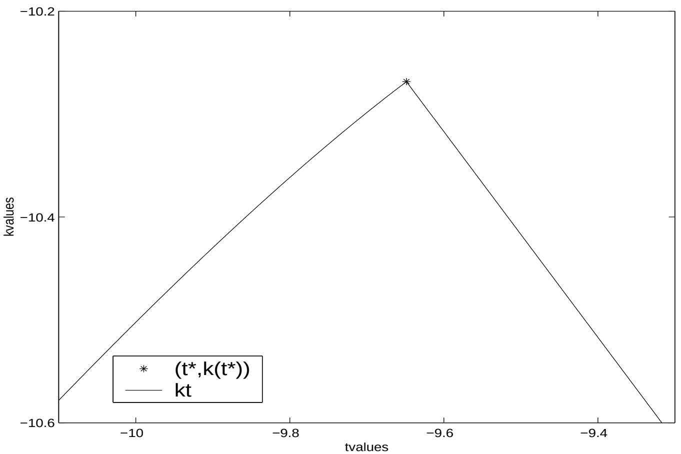


Figure: $k(t)$ in the hard case (case 2)

(Near) Hard Case $k(t) = (s^2 + 1)\lambda_1(D(t)) - t$

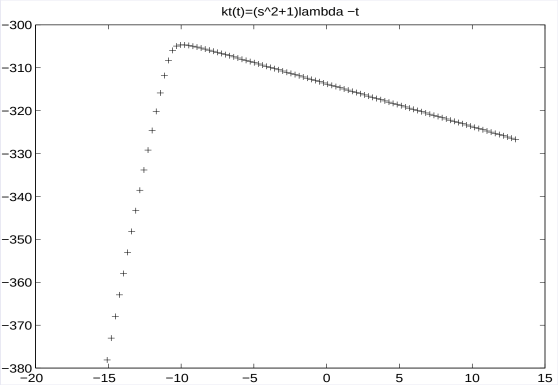


Figure: (near) hard case: $k(t) = (s^2 + 1)\lambda_1(D(t)) - t$

Hard Case: $k'(t) = (s^2+1)y_0^2(t) - 1$

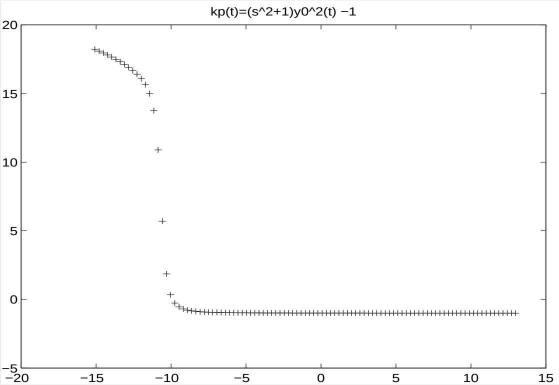


Figure: Hard Case: $k'(t) = (s^2+1)y_0^2(t) - 1$

Hard Case is Easiest Case: Shift and Deflation

If $\lambda_{\min} < 0$ use $A \leftarrow A - \lambda_{\min}(A)$

- hard case (case 2) happens only when $\lambda^* = \lambda_{\min}(A) < 0$
- $A \succeq 0$ after the shift (objective changes by a constant)

Lemma Fortin-W:03 [1]

$A = \sum_{i=1}^n \lambda_i(A) v_i v_i^T = Q \Lambda Q^T$ orthogonal spectral decomposition of A; $\gamma_i = (Q^T a)_i$

$$\begin{aligned}
S_1 &= \{i : \gamma_i \neq 0, \lambda_i(A) > \lambda_{\min}(A)\} \\
S_2 &= \{i : \gamma_i = 0, \lambda_i(A) > \lambda_{\min}(A)\} \\
S_3 &= \{i : \gamma_i \neq 0, \lambda_i(A) = \lambda_{\min}(A)\} \\
S_4 &= \{i : \gamma_i = 0, \lambda_i(A) = \lambda_{\min}(A)\}
\end{aligned}$$

For $k = 1, 2, 3, 4,$: $A_k = \sum_{i \in S_k} \lambda_i(A) v_i v_i^T$.

Then:

Lemma Conclusions

1. If $S_3 \neq \emptyset$ (easy case), then (x^*, λ^*) solves TRS <u>iff</u> (x^*, λ^*) solves TRS when A is replaced by $A_1 + A_3$.
2. If $S_3 = \emptyset$ (hard case), let $i_0 = 1 \in S_4$, then (x^*, λ^*) solves TRS <u>iff</u> (x^*, λ^*) solves TRS when A is replaced by $A_1 + \lambda_{i_0}(A) v_{i_0} v_{i_0}^T$. (DEFLATE)
3. Let $x(\lambda^*) = (A - \lambda^* I)^\dagger a$, then (x^*, λ^*), where $x^* = x(\lambda^*) + z, z \in \mathcal{N}(A - \lambda^* I)$ and $\|x^*\| = s$ solves TRS <u>iff</u> $(x(\lambda^*), \lambda^* - \lambda_{\min}(A))$ solves TRS when A is replaced by $A - \lambda_{\min}(A) I$.
4. If $\lambda_{\min}(A) \geq 0$, then (x^*, λ^*) solves TRS iff (x^*, λ^*) solves TRS when A is replaced by $A + \sum_{i \in S_4} \alpha_i v_i v_i^T$, with $\alpha_i \geq 0$. (DEFLATE all in S_4)

Shift/Deflate: Solve Hard Case Explicitly

Summary - sometimes you get !lucky!

- eigenpair $\lambda_{\min}(A) = \lambda_1(A), v_1$ found; large sparse case; so, assume possible hardcase: $\boxed{\lambda_1(A) < 0 \text{ and } v_1^T a = 0}$
- !we have $\lambda_1(A)$! so shift: $A \leftarrow A - \lambda_{\min}(A)I \succeq 0$; !optimum unchanged, objective value changed by $\lambda_1 s^2$!.
- !we have v_1, $v_1^T a = 0$! so deflate: $A \leftarrow A + \alpha_1 vv^T$, $\|A\| > \alpha_1 >> 0$; (repeat deflation if needed, i.e as long as $\boxed{\lambda_{\min}(A) = 0 \text{ and } v^T a = 0}$) !objective value/optima unchanged!.
- if $v^T a \neq 0$, then continue with TRS algorithm, i.e. hard case (case 1); otherwise, $A \succ 0$, so (using prec. conj grad.) calculate $\bar{x} = A^{-1}a$; if $\|x(\lambda^*)\| > s$, then continue with TRS algorithm, i.e. hard case (case 1); otherwise $\|x(\lambda^*)\| \leq s$, then we have hard case (case 2) and !we have an explicit solution : $\|x^*\| = \|\bar{x} + \beta v_1\| = s$, !since $v_1 \in \mathcal{N}(A - \lambda^* I)$!.

Techniques Used for Maximizing $k(t)$

"good side" and "bad side" for t

Let $t_0 := \lambda_{\min}(A) + \sum_{k \in \{j | (Q^T a)_j \neq 0\}} \frac{(Q^T a)_k^2}{\lambda_k(A) - \lambda_{\min}(A)}$

- If $k'(t) < 0$ ($t > t^*$), then t is on good side
- If $k'(t) > 0$ ($t < t^*$), then t is on bad side

Note: if $\exists t, t^* < t < t_0$ (on good side), then the hard case does not occur.

Techniques:

- Bracketing Newton's Method
- Triangle Interpolation
- Vertical Cut
- Inverse Linear Interpolation

First order method to maximize $k(t)$

duality provides upper and lower bounds of $k^* = \max_t k(t)$: k_{up}, k_{low}, respectively; $M_j := \alpha k_{up} + (1-\alpha)k_{low}$, $\alpha \in (0,1)$
$t_{j+1} = t_j - \frac{k(t_j)-M_j}{k'(t_j)}$: if $k(t_{j+1}) > k(t_j)$, then $k_{low} \leftarrow k(t_{j+1})$ (else $k_{up} \leftarrow M_j$)

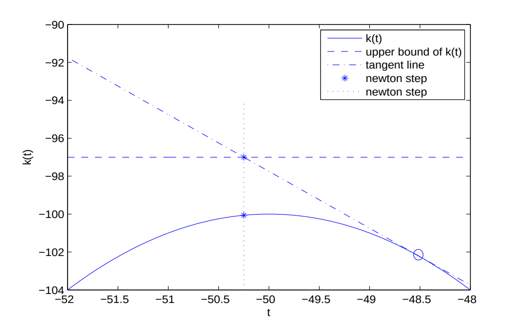


Figure: Bracketing Newton's Method step to solve $k(t) - M_j = 0$

Triangle Interpolation

Pick new t_+ if in current interval of uncertainty for t

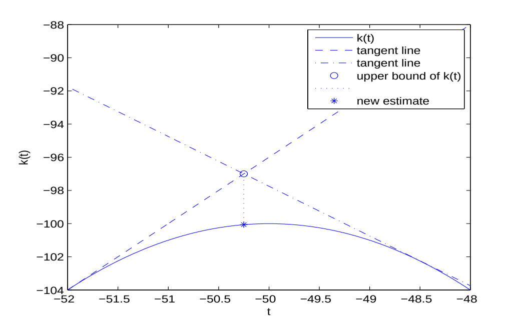


Figure: Triangle Interpolation

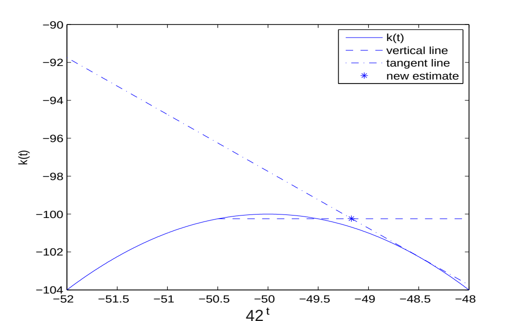
Vertical Cut

if $k'(t_g) < 0\ k'(t_b) > 0$, $k(t_g) < k(t_b)$, then:

use vertical cut to reduce the interval of the t values, i.e., solve

$$k(t_g) + k'(t_g)(t - t_g) = k(t_b)$$

the intersection of secant line on $(t_g, k(t_g))$ and the line $k = k(t_b)$ yields: the new t in (t_b, t_g) and on good side (Similarly if $k(t_b) < k(t_g)$)



Inverse Interpolation

$$\begin{bmatrix} \psi_1^2 & \psi_1 & 1 \\ \psi_2^2 & \psi_2 & 1 \\ \psi_3^2 & \psi_3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ t_{\text{new}} \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

on $\psi(t)$

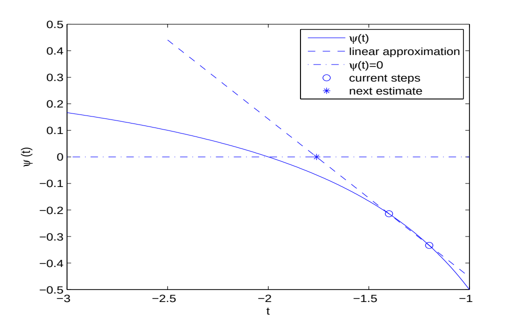


Figure: Linear Interpolation

Inverse interpolation; easy case & hard case (*case 1*)

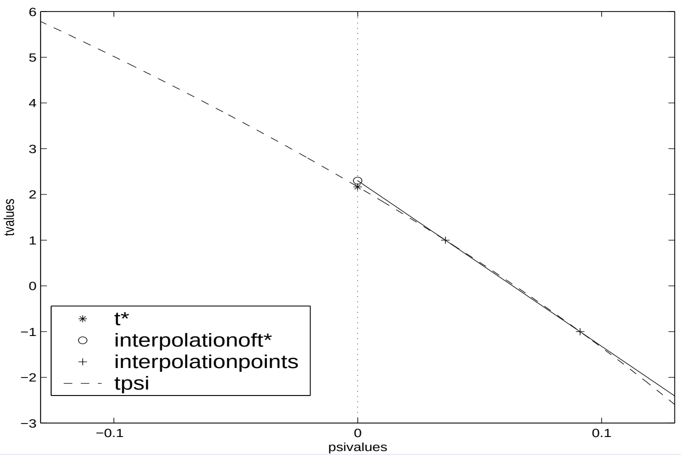


Figure: Inverse interpolation; easy case & hard case (*case 1*)

Inverse interpolation in the hard case (*case 2*)

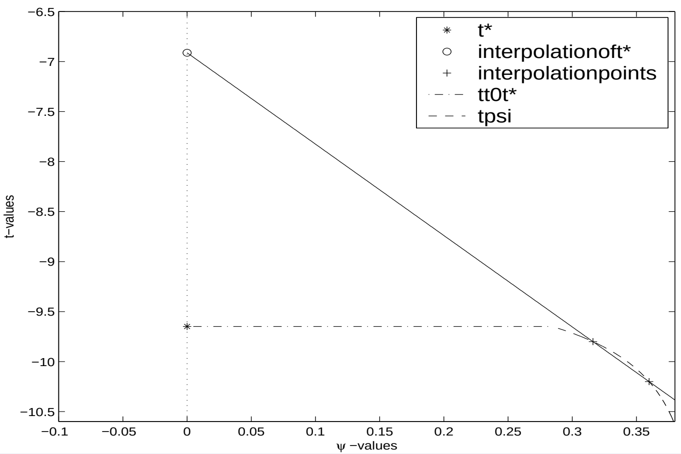


Figure: Inverse interpolation in the hard case (*case 2*)

$$\text{(TRS)} \qquad \begin{array}{rll} q^* = & \min & q(x) := \|Bx - b\|^2 \\ & \text{s.t.} & \|x\| = s \end{array}$$

- $A = B^T B \succeq 0, \operatorname{rank}(A) < n$; B is $m \times n, m > n$ and $\operatorname{rank}(B) < n$.
- $a = B^T b \in \mathcal{R}(A), a \perp \mathcal{N}(A - \lambda_{\min}(A)I)$
- so, hard case (case 2) if $\|A^\dagger a\| \leq s$.

Corollary: high accuracy explicit solution if s chosen large enough!

- deflate the eigenvectors corresponding to the 0 eigenvalues;
- use PCG to solve the system and use an eigenvector to move to the boundary.

cputime in the hard case (*case 2*)

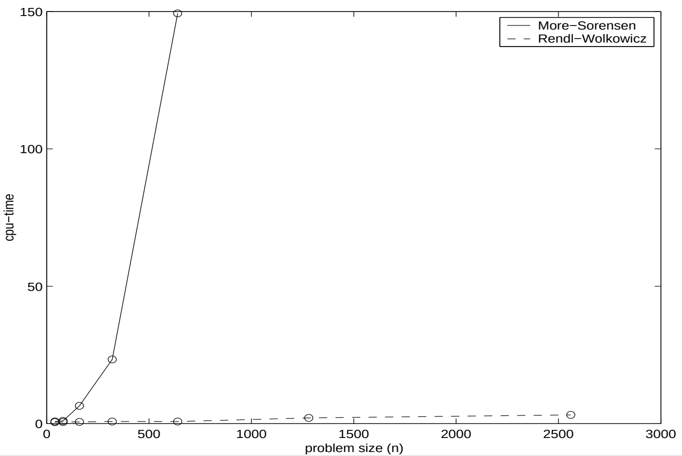


Figure: cputime in the hard case (*case 2*)

log of cputime in the hard case (*case 2*)

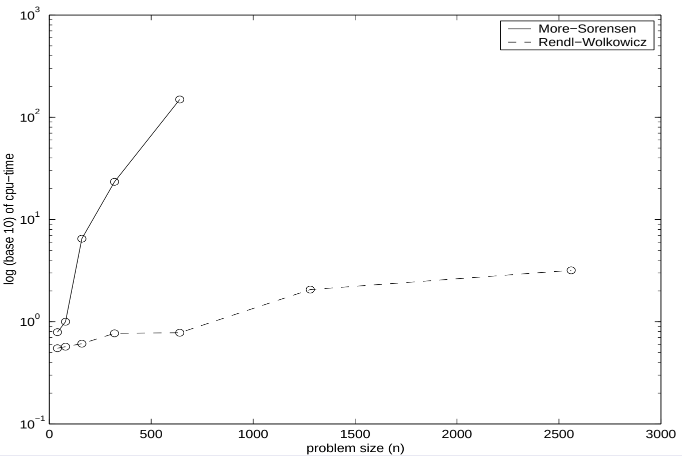


Figure: log of cputime in the hard case (*case 2*)

- C. Fortin and H. Wolkowicz, *The trust region subproblem and semidefinite programming*, Optim. Methods Softw. **19** (2004), no. 1, 41–67. MR MR2062235 (2005c:90065)

- D.M. Gay, *Computing optimal locally constrained steps*, SIAM J. Sci. Statist. Comput. **2** (1981), 186–197.

- M.D. Hebden, *An algorithm for minimization using exact second derivatives*, Tech. Report TP515, Atomic Energy Research Establishment, Harwell, England, 1973.

- Y. Levin and A. Ben-Israel, *The Newton bracketing method for convex minimization*, Computational Optimization and Applications (2001).

- J.J. Moré, *Generalizations of the trust region problem*, Optim. Methods Software **2** (1993), 189–209.

- J.J. Moré and D.C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput. **4** (1983), 553–572.

- I. Pólik and T. Terlaky, *A survey of the S-lemma*, SIAM Rev. **49** (2007), no. 3, 371–418 (electronic). MR MR2353804

- C. Reinsch, *Smoothing by spline functions*, Numer. Math. **10** (1967), 177–183.

- ______, *Smoothing by spline functions II*, Numer. Math. **16** (1971), 451–454.

- R. Stern and H. Wolkowicz, *Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations*, SIAM J. Optim. **5** (1995), no. 2, 286–313. MR 96h:90077

Efficient Solutions for the Large Scale Trust Region Subproblem, TRS

Henry Wolkowicz

(work with Ting Kei Pong and Heng Ye)
Dept. of Combinatorics and Optimization
University of Waterloo

WCOM FALL, 2011, at UBC Okanagan, Kelowna, BC