Efficient Solutions for Large Scale Trust Region Subproblem

Ting Kei Pong
Combinatorics & Optimization, University of Waterloo
Waterloo

13th Midwest Optimization Meeting
Oct 2011
(Ongoing work with Henry Wolkowicz, Heng Ye)

Outline

- Trust region subproblem and its properties.
- Easy case and hard cases.
- Rendl-Wolkowicz algorithm.
- Hard case: shift and deflate.
- Easy case: bracketing Newton's method...
- Numerical results.

Trust Region Subproblem (TRS)

$$q^* := \min q(x) := x^T A x - 2a^T x$$
$$\text{s.t. } \|x\| \leq s,$$

where $A \in \mathcal{S}^n$, $a \in \mathbb{R}^n$, $s > 0$.

Applications: TR methods for unconstr. min., subproblems for constrained optimization, regularization of ill-posed problems...

Trust Region Subproblem (TRS)

$$q^* := \min q(x) := x^T Ax - 2a^T x$$
$$\text{s.t. } \|x\| \le s,$$

where $A \in \mathcal{S}^n$, $a \in \mathbb{R}^n$, $s > 0$.

Applications: TR methods for unconstr. min., subproblems for constrained optimization, regularization of ill-posed problems...

Fact 1 (Gay '81; More, Sorensen '83): x^* is optimal for TRS iff $\exists \lambda^*$ s.t.

$$\left.\begin{array}{c}(A - \lambda^* I)x^* = a, \\ A - \lambda^* I \succeq 0, \lambda^* \le 0,\end{array}\right\} \quad \text{dual feasibility}$$
$$\|x^*\|^2 \le s^2, \quad \text{primal feasibility}$$
$$\lambda^*(s^2 - \|x^*\|^2) = 0. \quad \text{complementary slackness}$$

MoSo Algorithm Framework for TRS

Suppose $A - \lambda^* I \succ 0$.

- Define $x(\lambda) = (A - \lambda I)^{-1} a$.
- Solve $\psi(\lambda) := \|x(\lambda)\|^2 - s^2 = 0$.
- Maintain $A - \lambda I \succ 0$, $\lambda \leq 0$.

Remarks:

- Solve less nonlinear $\phi(\lambda) := \frac{1}{s} - \frac{1}{\|x(\lambda)\|} = 0$ (Reinsch '67; Hebden '73).
- Each iteration involves a Cholesky decomposition of $A - \lambda^k I$.

Easy/Hard Cases for TRS

Let $A = Q\Lambda Q^T$ be eigenvalue decomposition; $\gamma = Q^T a$.

$$\psi(\lambda) = \|x(\lambda)\|^2 - s^2 = \sum_{j=1}^{n} \frac{\gamma_j^2}{(\lambda_j(A) - \lambda)^2} - s^2.$$

Easy case	Hard case 1	Hard case 2
$a \notin \mathcal{R}(A - \lambda_{\min}(A)I)$ $(\Rightarrow \lambda^* < \lambda_{\min}(A))$	$a \perp \mathcal{N}(A - \lambda_{\min}(A)I)$ but $\lambda^* < \lambda_{\min}(A)$	$a \perp \mathcal{N}(A - \lambda_{\min}(A)I)$ and $\lambda^* = \lambda_{\min}(A)$ (i) $\|(A - \lambda^* I)^\dagger a\| = s$ or $\lambda^* = 0$ (ii) $\|(A - \lambda^* I)^\dagger a\| < s, \lambda^* < 0$

Rendl-Wolkowicz Algorithm

Consider equality constrained problem $\mu^* = \min_{||x||=s} x^TAx - 2a^Tx$.

$$\begin{aligned}
\mu^* &= \min_{||x||=s,\ y_0^2=1} x^TAx - 2y_0a^Tx \\
&= \max_{t} \min_{||x||=s,\ y_0^2=1} x^TAx - 2y_0a^Tx + ty_0^2 - t \\
&\geq \max_{t} \min_{||x||^2+y_0^2=s^2+1} x^TAx - 2y_0a^Tx + ty_0^2 - t \quad \color{red}{**\text{ eig prob } **} \\
&= \max_{t,\lambda} \min_{x,y_0} x^TAx - 2y_0a^Tx + ty_0^2 - t + \lambda(||x||^2 + y_0^2 - s^2 - 1) \\
&= \max_{r,\lambda} \min_{x,y_0} x^TAx - 2y_0a^Tx + ry_0^2 - r + \lambda(||x||^2 - s^2) \\
&= \max_{\lambda} \min_{x,y_0^2=1} x^TAx - 2y_0a^Tx + \lambda(||x||^2 - s^2) = \mu^*.
\end{aligned}$$

Rendl-Wolkowicz Algorithm

From the eig prob:

$$\mu^* = \max_t \underbrace{\min_{||x||^2+y_0^2=s^2+1} x^T A x - 2y_0 a^T x + t y_0^2}_{(s^2+1)\lambda_{\min}(D(t))} - t$$

$$= \max_t k(t) := (s^2+1)\lambda_{\min}(D(t)) - t$$

where

$$D(t) = \begin{pmatrix} t & -a^T \\ -a & A \end{pmatrix}.$$

Observe that the candidate for interior solution is $A^{-1}a$, when $A \succ 0$.

Let t^* be the global maximizer of $k(t)$.

Observe that the candidate for interior solution is $A^{-1}a$, when $A \succ 0$.

Let t^* be the global maximizer of $k(t)$.

Fact 2 (Rendl, Wolkowicz '97): If $\lambda_{\min}(D(t^*))$ is simple, $y(t^*) = \begin{pmatrix} y_0(t^*) \\ w(t^*) \end{pmatrix}$ is the normalized eigenvector for $\lambda_{\min}(D(t^*))$, then $y_0(t^*) \neq 0$ and

$$\frac{1}{|y_0(t^*)|}\|w(t^*)\| = s.$$

Observe that the candidate for interior solution is $A^{-1}a$, when $A \succ 0$.

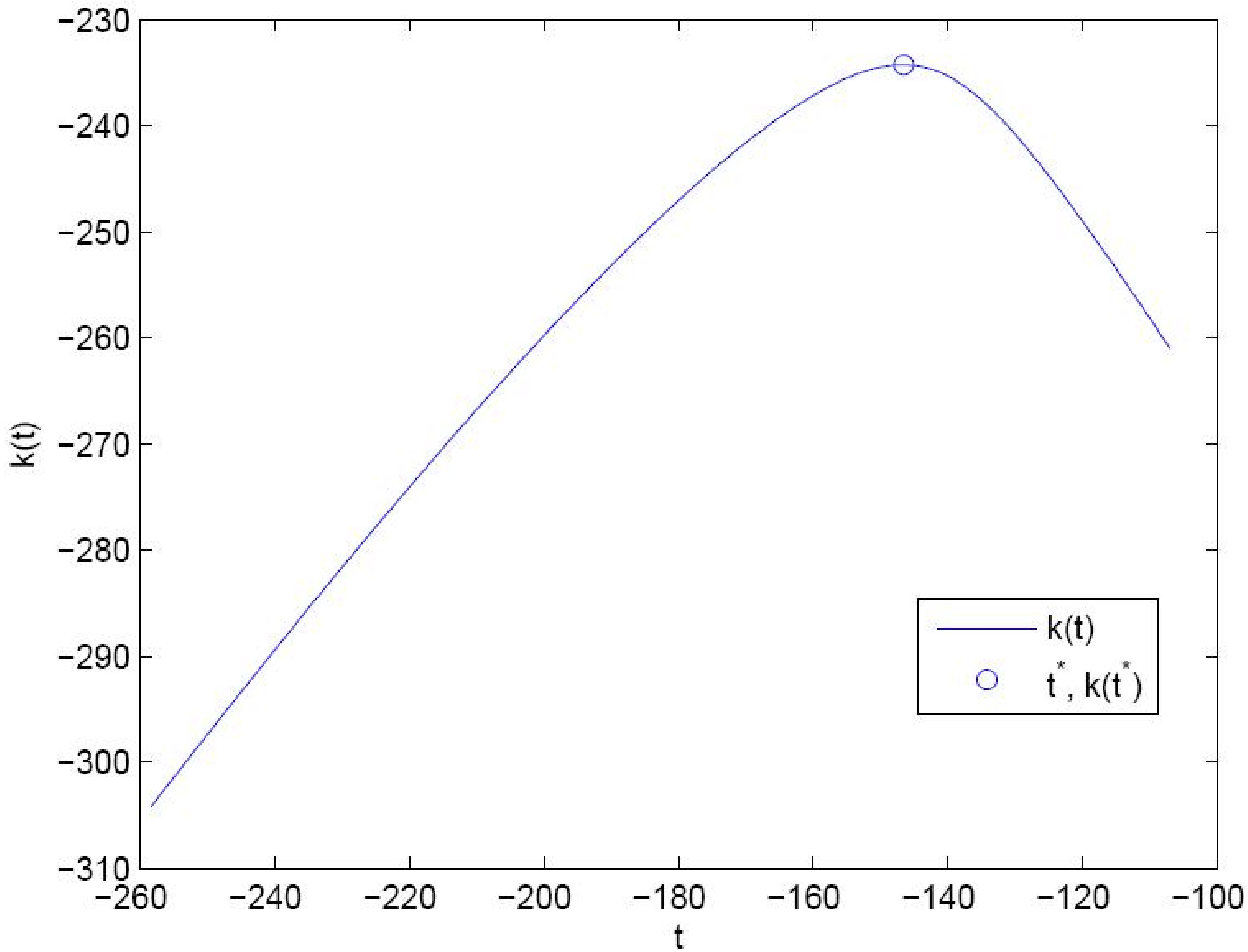
Let t^* be the global maximizer of $k(t)$.

Fact 2 (Rendl, Wolkowicz '97): If $\lambda_{\min}(D(t^*))$ is simple, $y(t^*) = \begin{pmatrix} y_0(t^*) \\ w(t^*) \end{pmatrix}$ is the normalized eigenvector for $\lambda_{\min}(D(t^*))$, then $y_0(t^*) \neq 0$ and

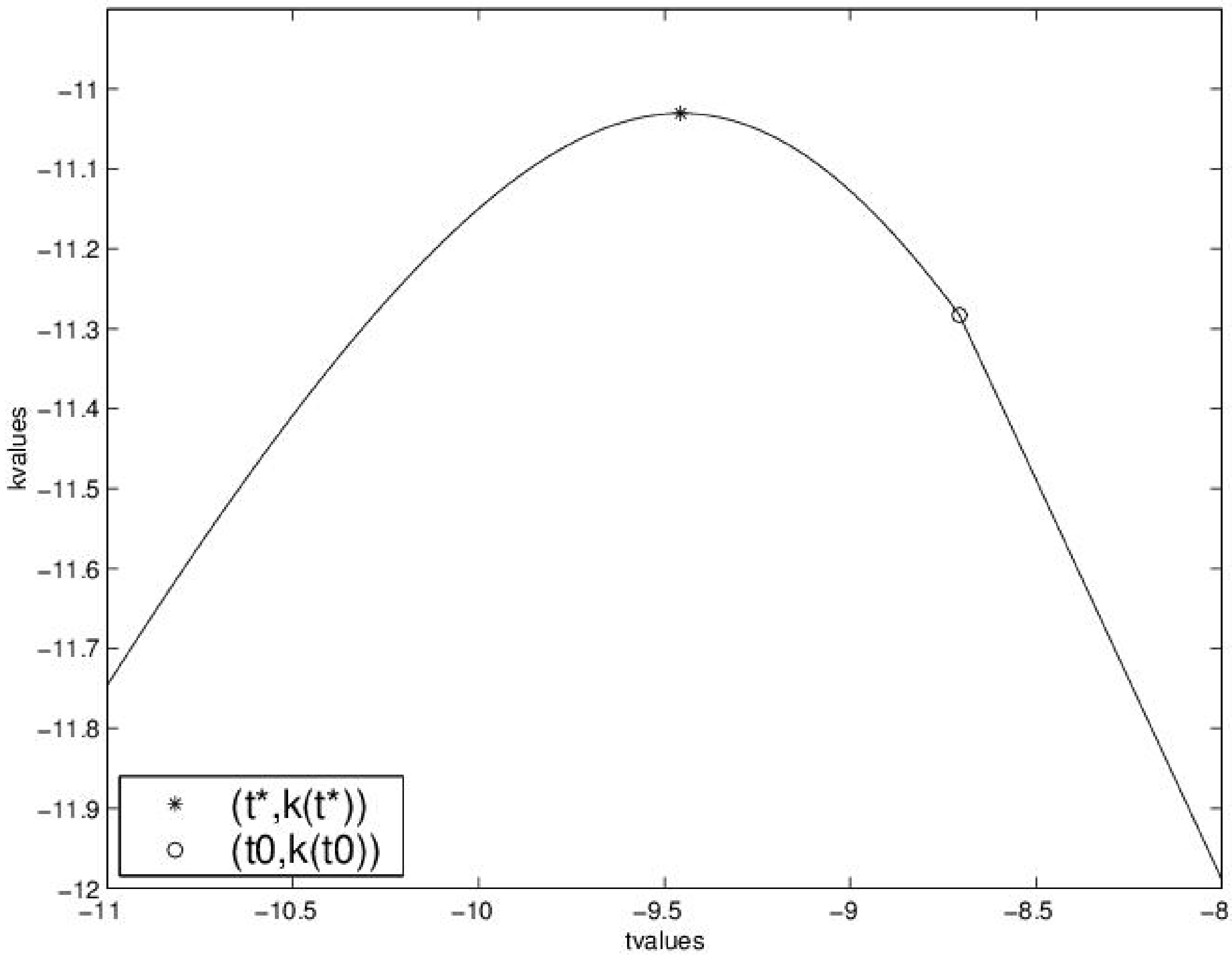
$$\frac{1}{|y_0(t^*)|}\|w(t^*)\| = s.$$

Note: $\lambda_{\min}(D(t^*))$ is simple for easy case and hard case 1.

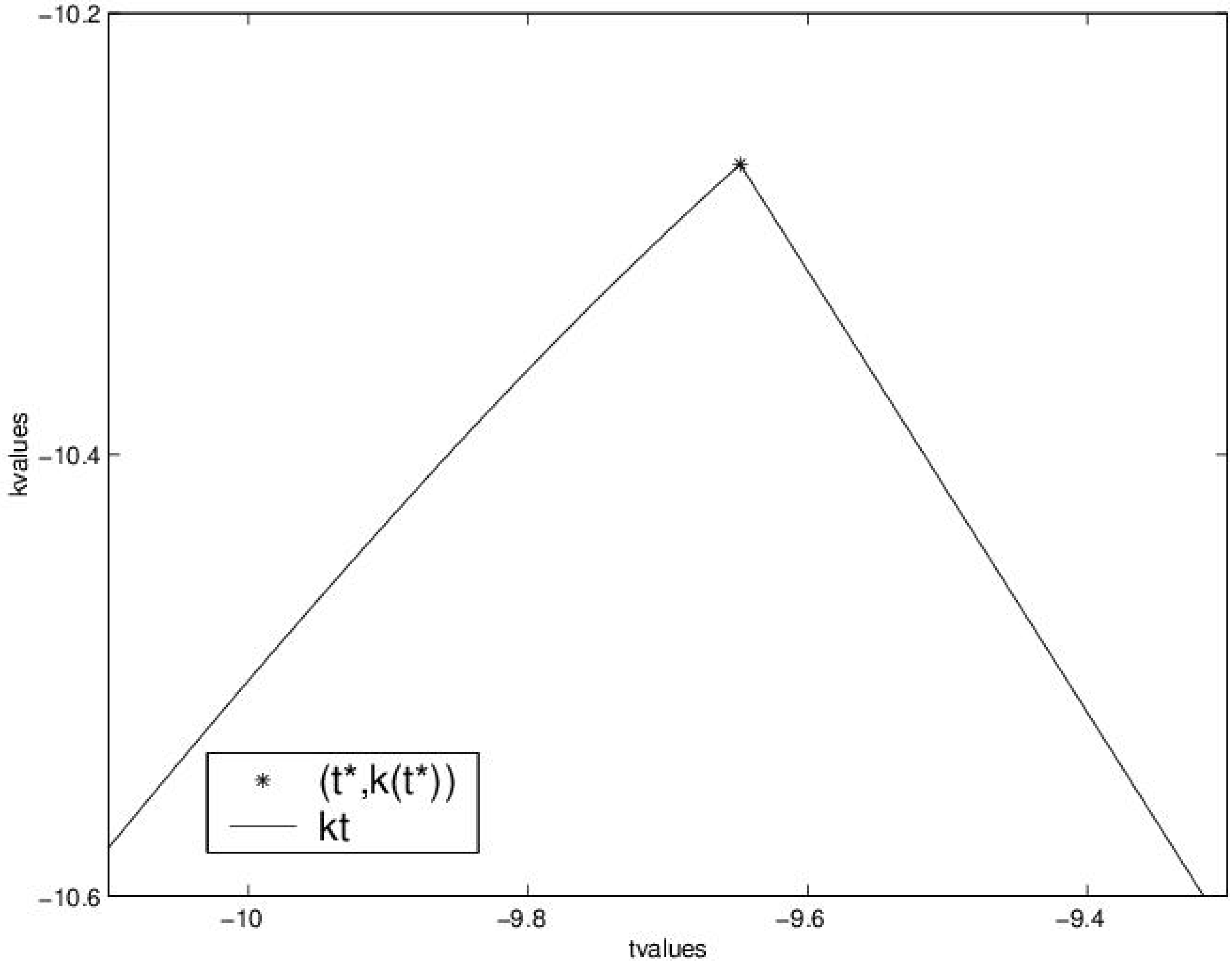
$k(t)$ in Easy Case



$k(t)$ in Hard Case 1



$k(t)$ in Hard Case 2



Hard Case is Easiest: Shift and Deflate

Fact 3 (Fortin, Wolkowicz '03): Let $A = \sum_{i=1}^{n} \lambda_i(A) v_i v_i^T = Q \Lambda Q^T$ be the orthogonal spectral decomposition of A; $\gamma_i = (Q^T a)_i$

$$\begin{array}{rcl} S_1 & = & \{i : \gamma_i \neq 0, \lambda_i(A) = \lambda_{\min}(A)\} \\ S_2 & = & \{i : \gamma_i = 0, \lambda_i(A) = \lambda_{\min}(A)\} \end{array}$$

- Let $x(\lambda^*) = (A - \lambda^* I)^\dagger a$, then $(x(\lambda^*), \lambda^* - \lambda_{\min}(A))$ solves TRS with $A - \lambda_{\min}(A)I$ in place of $A \Leftrightarrow$ (x^*, λ^*) solves TRS, where $x^* = x(\lambda^*) + z$, $z \in \mathcal{N}(A - \lambda^* I)$ and $\|x^*\| = s$.

- If $\lambda_{\min}(A) \geq 0$, then (x^*, λ^*) solves TRS $\Leftrightarrow$ (x^*, λ^*) solves TRS when A is replaced by $A + \sum_{i \in S_2} \alpha_i v_i v_i^T$, with $\alpha_i \geq 0$.

Solving Hard Case 2 Explicitly

- Use Lanczos, find $\lambda_{\min}(A), v_1$; assume possible hard case:
  $\boxed{\lambda_{\min}(A) < 0 \text{ and } v_1^T a = 0}$;

- Shift: $A \leftarrow A - \lambda_{\min}(A)I \succeq 0$;

- Deflate: $A \leftarrow A + \alpha_1 v v^T$, $\|A\| > \alpha_1 >> 0$. Repeat deflation as long as
  $\boxed{\lambda_{\min}(A) = 0 \text{ and } v^T a = 0}$;

- If $v^T a \neq 0$, we are in easy case; otherwise, $A \succ 0$, calculate $\bar{x} = A^{-1}a$ using prec. conj grad.
  If $\|x(\lambda^*)\| > s$, we are in hard case 1; otherwise $\|x(\lambda^*)\| \leq s$, then we have an explicit solution:

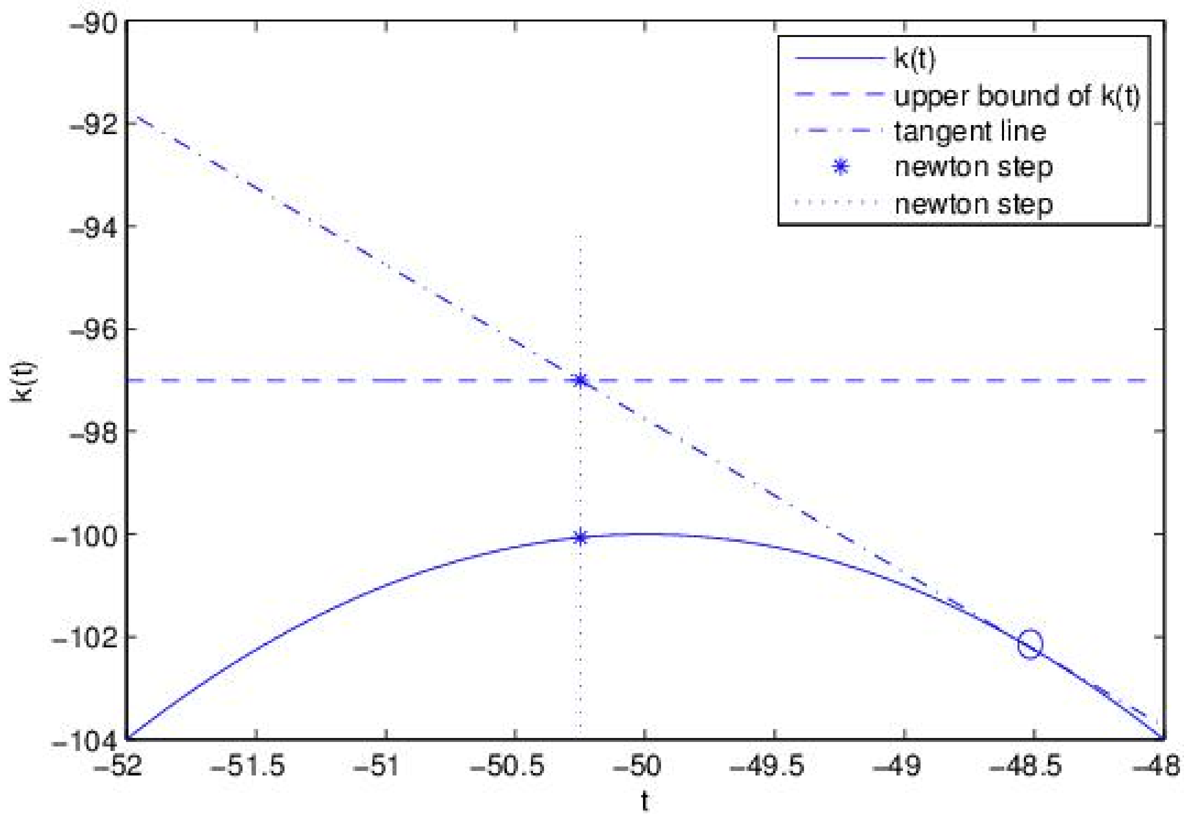
$$\|x^*\| = \|x(\lambda^*) + \beta v_1\| = s, \qquad v_1 \in \mathcal{N}(A_{\text{orig}} - \lambda^* I).$$

Maximizing $k(t)$

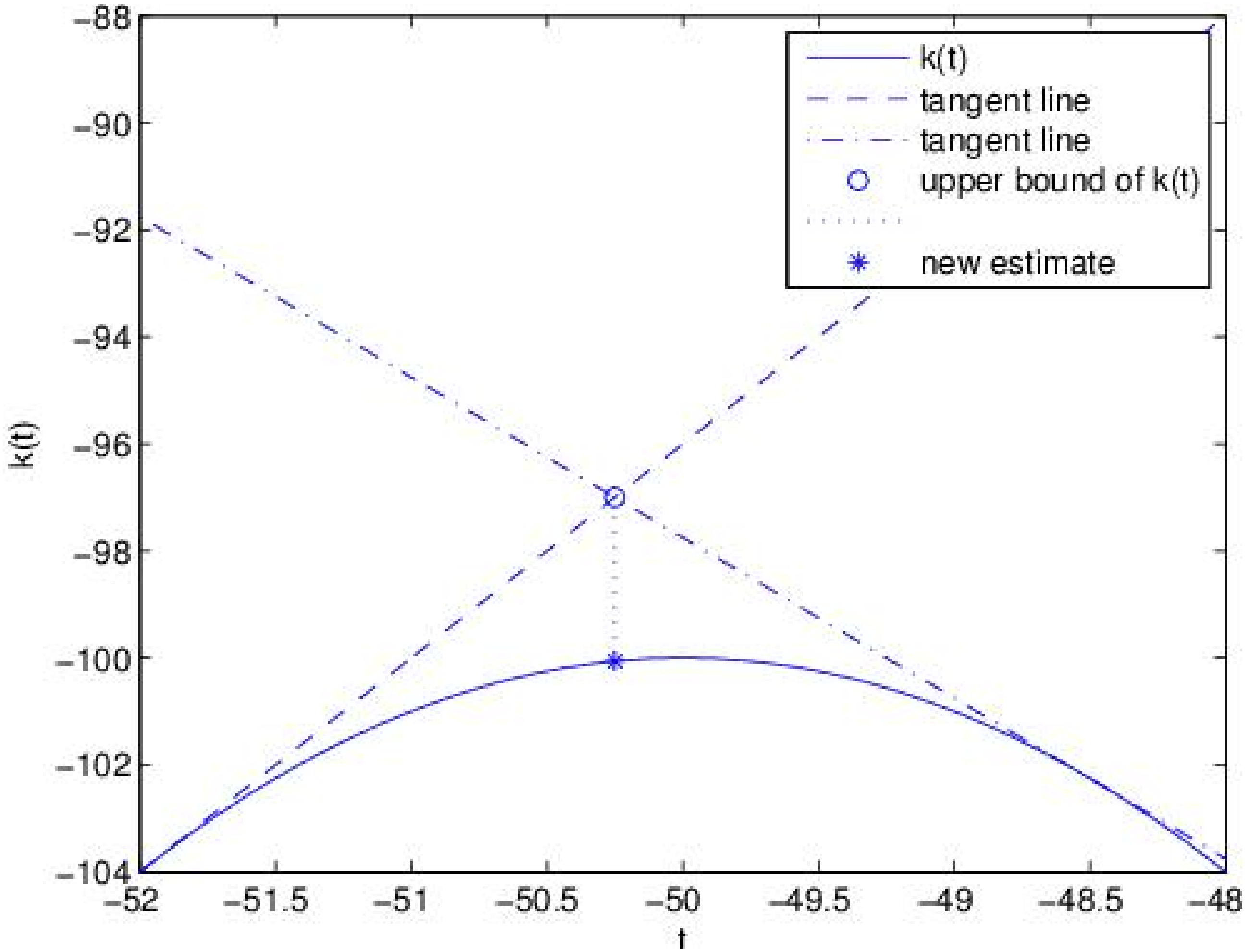
- Bracketing Newton's Method (Ben-Israel ,Levin '01).
- Triangle Interpolation.
- Vertical Cut.
- Inverse Linear Interpolation.

Bracketing Newton

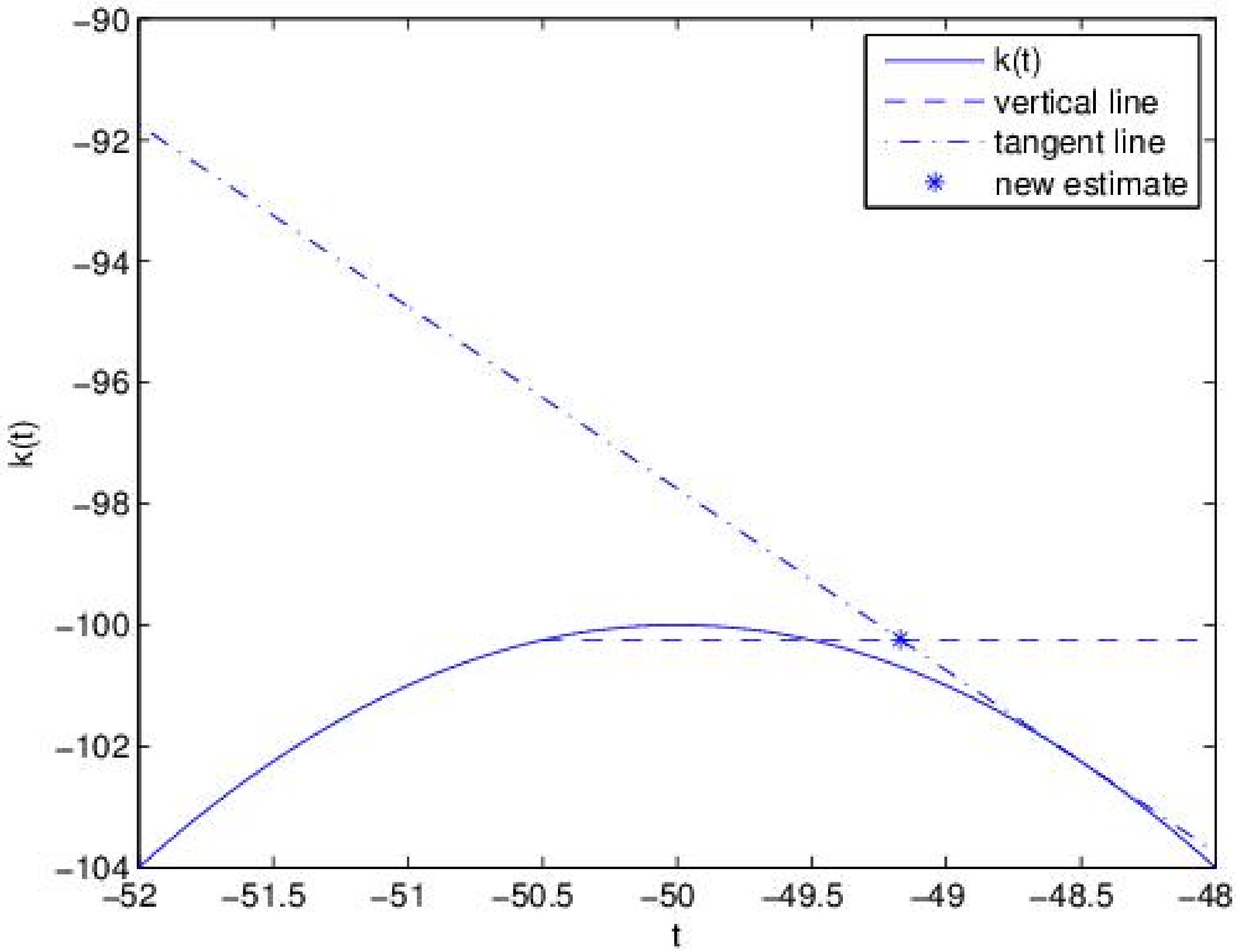
Takes one Newton step for solving $k(t) = M_j$, where M_j is an estimate of the optimal value from previous iterates.



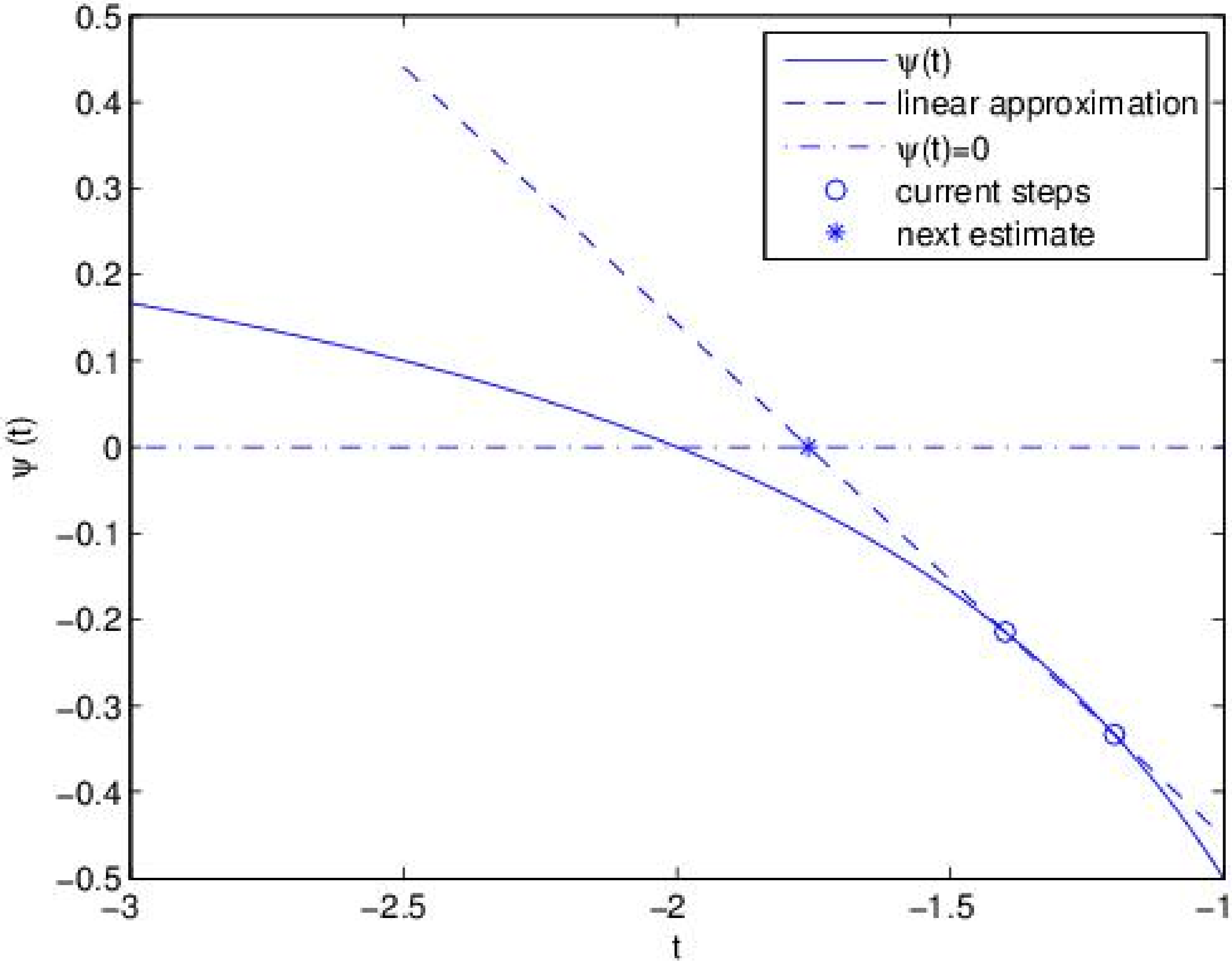
Triangle Interpolation



Vertical Cut



Inverse Linear Interpolation

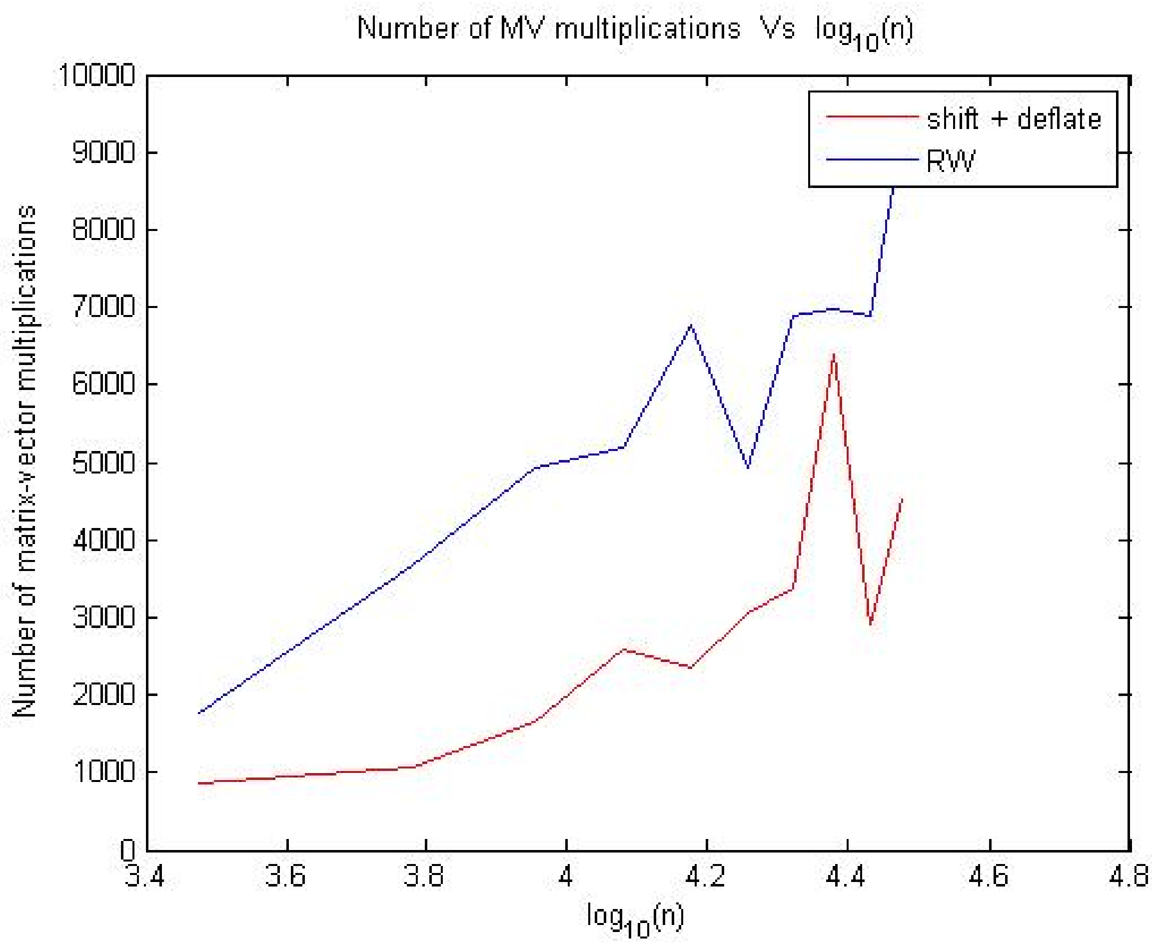


Simulations

- Compare RW with and without shift+deflate on hard case 2 instances.
- For $n = 3000, 6000, ..., 30000$, generate 10 hard instances:

```
A = sprandsym(n,0.01);
[v,lambda] = eigs(A,1,'SA',opts);
xtempopt = randn(n,1);
s = 1.1*norm(xtempopt);
a = A*xtempopt - lambda*xtempopt;
```

Simulations



Conclusion & Future work

- After shift and deflation, hard case becomes easy.
- $k(t)$ can be maximized efficiently by simple techniques.
- More numerical tests...
- Fast algorithm for solving Generalized TRS:

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

Thanks for coming! :)