# Generalized Trust Region Subproblem: Analysis and Algorithm

Ting Kei Pong
Combinatorics & Optimization, University of Waterloo
Waterloo

ISMP 2012
Aug 2012
(Ongoing work with Henry Wolkowicz)

# Outline

- Generalized trust region subproblem.
- Optimality conditions.
- Easy case and hard cases.
- (Extended) Rendl-Wolkowicz algorithm.
- Preliminary numerical results.

# Generalized Trust Region Subproblem (GTRS)

The generalized trust region subproblem:

$$\begin{aligned} q^* := \min\ q(x) &:= x^T A x - 2a^T x \\ \text{s.t. } l \leq &\underbrace{x^T B x - 2b^T x}_{q_1(x)} \leq u, \end{aligned}$$

where $A, B \in \mathcal{S}^n$, $a, b \in \mathbb{R}^n$.

- Reduces to trust region subproblem (TRS) when $B = I$, $b = 0$, $0 = l < u$.
- Possible applications: TR methods for unconstr. min., subproblems for constrained optimization, regularization of ill-posed problems...

# Assumptions on GTRS

- $B \neq 0$
- GTRS is feasible and bounded below.
- The following constraint qualification holds

$$\text{(CQ)} \qquad \operatorname{tr}(B\hat{X}) - 2b^T\hat{x} \in \operatorname{ri}([\ell, u]), \qquad \text{for some } \hat{X} \succ \hat{x}\hat{x}^T.$$

- The dual problem is feasible; i.e.,

$$\exists\lambda \text{ s.t. } a - \lambda b \in \mathcal{R}(A - \lambda B) \text{ and } A - \lambda B \succeq 0.$$

# Optimality conditions

**Fact 1** (P, Wolkowicz '12): $x^*$ optimal for GTRS iff $\exists\lambda^*$ s.t.

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

$$\ell \le {x^*}^T B x^* - 2b^T x^* \le u, \quad \text{primal feasibility}$$

$$\left.\begin{array}{c}(\lambda^*)_+(\ell - {x^*}^T B x^* + 2b^T x^*) = 0,\\ (\lambda^*)_-({x^*}^T B x^* - 2b^T x^* - u) = 0.\end{array}\right\} \quad \text{complementary slackness}$$

# Optimality conditions

**Fact 1** (P, Wolkowicz '12): $x^*$ optimal for GTRS iff $\exists \lambda^*$ s.t.

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

$$\ell \leq {x^*}^T B x^* - 2b^T x^* \leq u, \quad \text{primal feasibility}$$

$$\left.\begin{array}{r}(\lambda^*)_+(\ell - {x^*}^T B x^* + 2b^T x^*) = 0,\\ (\lambda^*)_-({x^*}^T B x^* - 2b^T x^* - u) = 0.\end{array}\right\} \quad \text{complementary slackness}$$

**Fact 2** (Gay '81; More, Sorensen '83): $x^*$ is optimal for TRS iff $\exists \lambda^*$ s.t.

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

$$\|x^*\|^2 \leq u, \quad \text{primal feasibility}$$

$$\lambda^*(u - \|x^*\|^2) = 0. \quad \text{complementary slackness}$$

# Further Assumptions on GTRS

Assume $\exists \hat{\lambda}$ s.t.

$$A - \hat{\lambda} B \succ 0.$$

# Further Assumptions on GTRS

Assume $\exists \hat{\lambda}$ s.t.

$$A - \hat{\lambda} B \succ 0.$$

Then:

- GTRS has an optimal solution.
- There exists invertible matrix $S$ s.t.

$$A = S\mathrm{diag}(\alpha)S^T \quad B = S\mathrm{diag}(\beta)S^T.$$

- $$A - \lambda B \succeq 0 \;\Leftrightarrow \underline{\lambda} := \max_{\{i:\beta_i<0\}} \frac{\alpha_i}{\beta_i} \leq \lambda \leq \min_{\{i:\beta_i>0\}} \frac{\alpha_i}{\beta_i} =: \overline{\lambda}.$$

- For simplicity, we discuss only $s := u = l$.

# Easy/Hard Cases for GTRS

Consider $\psi(\lambda) = q_1(x(\lambda)) - s$, with $x(\lambda) := (A - \lambda B)^{\dagger}(a - \lambda b)$, $\underline{\lambda} \le \lambda \le \overline{\lambda}$.

| | Easy case | Hard case 1 | Hard case 2 |
|---|---|---|---|
| $\underline{\lambda}, \overline{\lambda}$ finite | $a - \overline{\lambda} b \notin \mathcal{R}(A - \overline{\lambda} B)$<br>and<br>$a - \underline{\lambda} b \notin \mathcal{R}(A - \underline{\lambda} B)$<br><br>$(\Rightarrow \underline{\lambda} < \lambda^* < \overline{\lambda})$ | $a - \overline{\lambda} b \perp \mathcal{N}(A - \overline{\lambda} B)$<br>or<br>$a - \underline{\lambda} b \perp \mathcal{N}(A - \underline{\lambda} B)$<br><br>$\underline{\lambda} < \lambda^* < \overline{\lambda}$ | $a - \overline{\lambda} b \perp \mathcal{N}(A - \overline{\lambda} B)$<br>$\lambda^* = \overline{\lambda}$<br>or<br>$a - \underline{\lambda} b \perp \mathcal{N}(A - \underline{\lambda} B)$<br>$\lambda^* = \underline{\lambda}$ |
| $\underline{\lambda} = -\infty$ | $a - \overline{\lambda} b \notin \mathcal{R}(A - \overline{\lambda} B)$<br><br>$(\Rightarrow -\infty < \lambda^* < \overline{\lambda})$ | $a - \overline{\lambda} b \perp \mathcal{N}(A - \overline{\lambda} B)$<br><br>$-\infty < \lambda^* < \overline{\lambda}$ | $a - \overline{\lambda} b \perp \mathcal{N}(A - \overline{\lambda} B)$<br><br>$\lambda^* = \overline{\lambda}$ |
| $\overline{\lambda} = \infty$ | $a - \underline{\lambda} b \notin \mathcal{R}(A - \underline{\lambda} B)$<br><br>$(\Rightarrow \underline{\lambda} < \lambda^* < \infty)$ | $a - \underline{\lambda} b \perp \mathcal{N}(A - \underline{\lambda} B)$<br><br>$\underline{\lambda} < \lambda^* < \infty$ | $a - \underline{\lambda} b \perp \mathcal{N}(A - \underline{\lambda} B)$<br><br>$\lambda^* = \underline{\lambda}$ |

# Hard Case 2: Explicit Solution

**Fact 3** (P, Wolkowicz '12): If $\overline{\lambda}$ is finite, then

$$v^T Bv > 0, \ \ \forall v \in \mathcal{N}(A - \overline{\lambda} B) \backslash \{0\}.$$

Similarly, if $\underline{\lambda}$ is finite, then

$$v^T Bv < 0, \ \ \forall v \in \mathcal{N}(A - \underline{\lambda} B) \backslash \{0\}.$$

# Hard Case 2: Explicit Solution

**Fact 3** (P, Wolkowicz '12): If $\overline{\lambda}$ is finite, then

$$v^T B v > 0, \quad \forall v \in \mathcal{N}(A - \overline{\lambda} B) \backslash \{0\}.$$

Similarly, if $\underline{\lambda}$ is finite, then

$$v^T B v < 0, \quad \forall v \in \mathcal{N}(A - \underline{\lambda} B) \backslash \{0\}.$$

Since $\psi$ is non-decreasing on $(\underline{\lambda}, \overline{\lambda})$,

- When $\lambda^* = \overline{\lambda}$, the solution is $x(\overline{\lambda}) + v$ for some $v \in \mathcal{N}(A - \overline{\lambda} B)$;
- When $\lambda^* = \underline{\lambda}$, the solution is $x(\underline{\lambda}) + v$ for some $v \in \mathcal{N}(A - \underline{\lambda} B)$.

# Extended Rendl-Wolkowicz Algorithm

Consider equality constrained problem $\mu^* = \min_{x^TBx-2b^Tx=s} x^TAx - 2a^Tx$ and assume $s \neq -1$. Then

$$\mu^* = \max_t k(t) := k_0(t) - t$$

where

$$\begin{aligned} k_0(t) &:= \inf_{x^TBx-2b^Txy_0+y_0^2=s+1} x^TAx - 2a^Txy_0 + ty_0^2 \\ &= \inf\left\{ y^T \begin{bmatrix} t & -a^T \\ -a & A \end{bmatrix} y : \ y^T \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} y = s+1 \right\}. \end{aligned}$$

$$k_0(t) = \inf\left\{ y^T \begin{bmatrix} t & -a^T \\ -a & A \end{bmatrix} y :\ y^T \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} y = s+1 \right\}.$$

- In TRS, $B = I$, $b = 0$ and $s > 0$, evaluating $k_0(t) \Leftrightarrow$ finding min. eig.;

- When $\begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} \succ 0$, evaluating $k_0(t) \Leftrightarrow$ finding min. gen. eig.;

$$k_0(t) = \inf \left\{ y^T \begin{bmatrix} t & -a^T \\ -a & A \end{bmatrix} y : \; y^T \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} y = s+1 \right\}.$$

- In TRS, $B = I$, $b = 0$ and $s > 0$, evaluating $k_0(t) \Leftrightarrow$ finding min. eig.;
- When $\begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} \succ 0$, evaluating $k_0(t) \Leftrightarrow$ finding min. gen. eig.;
- In general, using a $\hat{\lambda}$ with

$$\tilde{A} := \begin{bmatrix} t & -a^T \\ -a & A \end{bmatrix} - \hat{\lambda} \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} \succ 0,$$

evaluating $k_0(t) \Leftrightarrow$ finding min. gen. eig.

Suppose $s + 1 \neq 0$.

Suppose $s + 1 \neq 0$.

**Fact 4** (P, Wolkowicz '12): Suppose that $B$ is either indefinite or positive semidefinite. Let

$$t_- := -\inf_{x^T Bx - 2b^T x = -1} x^T Ax - 2a^T x,$$
$$s_- := \inf x^T Bx - 2b^T x.$$

- If $s_- < -1$, then $\mathrm{dom}(k) = [t_-, \infty)$;
- If $s_- > -1$, then $k(t)$ is defined everywhere.

In either case, $k$ is continuous and concave on its domain, and is differentiable in the interior of the domain except possibly at one point.

Suppose $s + 1 \neq 0$ and $t^*$ be the global maximizer of $k(t)$.

Suppose $s+1 \neq 0$ and $t^*$ be the global maximizer of $k(t)$.

**Fact 5** (P, Wolkowicz '12): In the easy case/hard case 1, $t^* > t_-$ and there exists a generalized eigenvalue $\lambda^* \in (\underline{\lambda}, \overline{\lambda})$ of the matrix pencil

$$\left( \begin{bmatrix} t^* & -a^T \\ -a & A \end{bmatrix}, \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} \right)$$

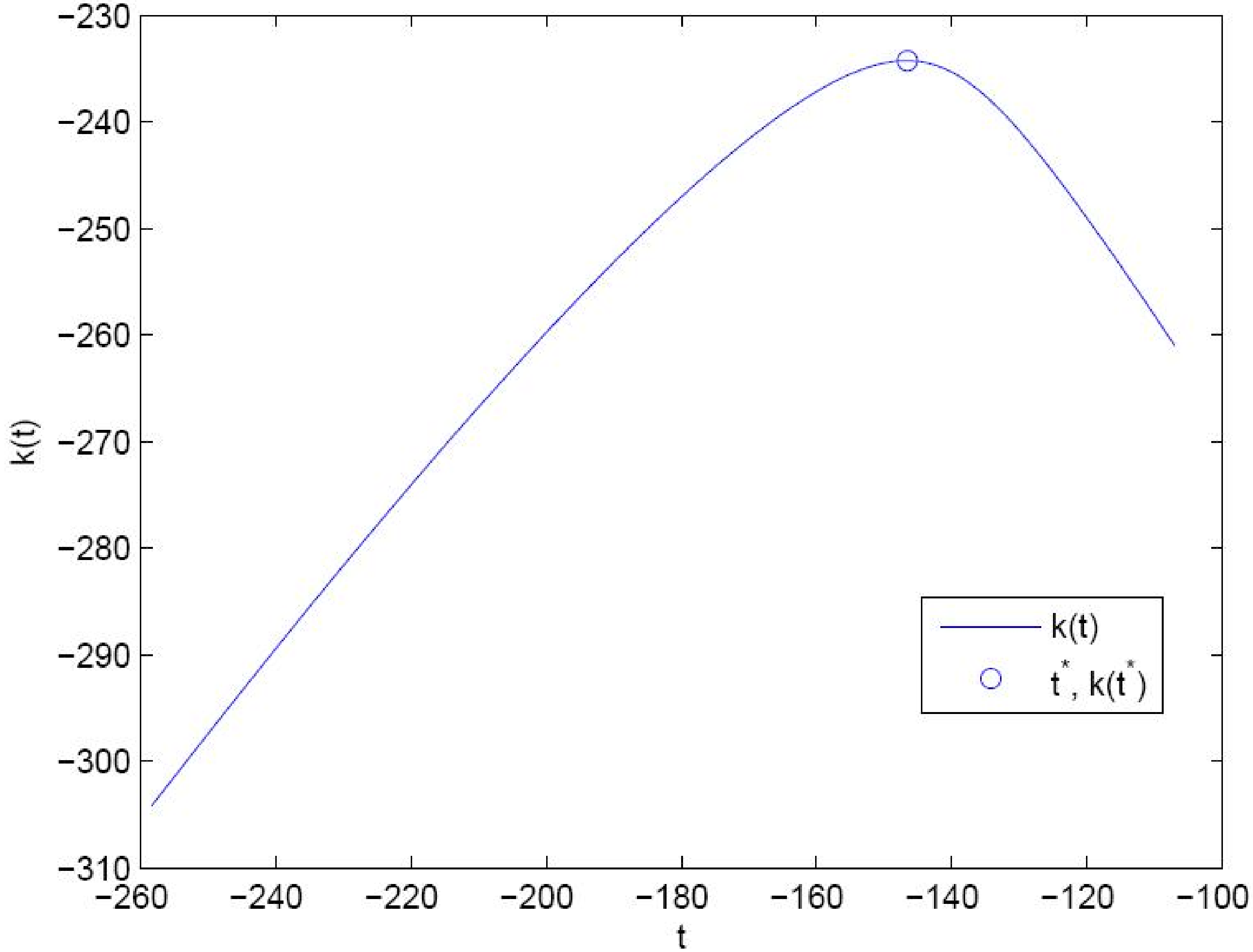
with a generalized eigenvector $y(t^*) = \begin{bmatrix} y_0(t^*) \\ w(t^*) \end{bmatrix}$ satisfying

$$y_0(t^*) \neq 0 \text{ and } y(t^*)^T \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} y(t^*) = \operatorname{sign}(s+1).$$

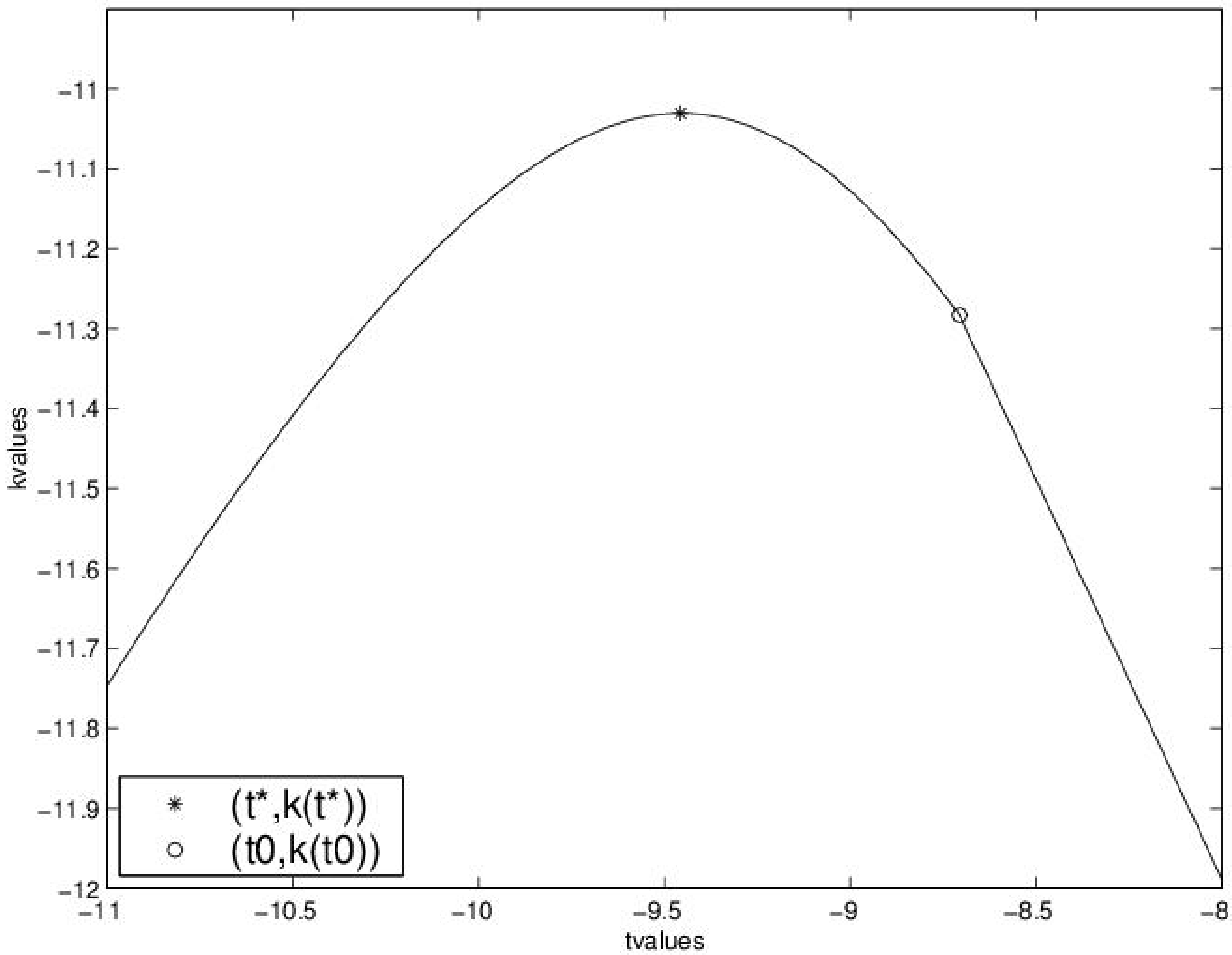
Moreover, $x^* := \frac{w(t^*)}{y_0(t^*)}$ solves GTRS, i.e.,

$$x^* = \underset{x^TBx - 2b^Tx = s}{\arg\min} \; x^TAx - 2a^Tx.$$

# $k(t)$ in Easy Case, $s_- > -1$



# $k(t)$ in Hard Case 1, $s_- > -1$



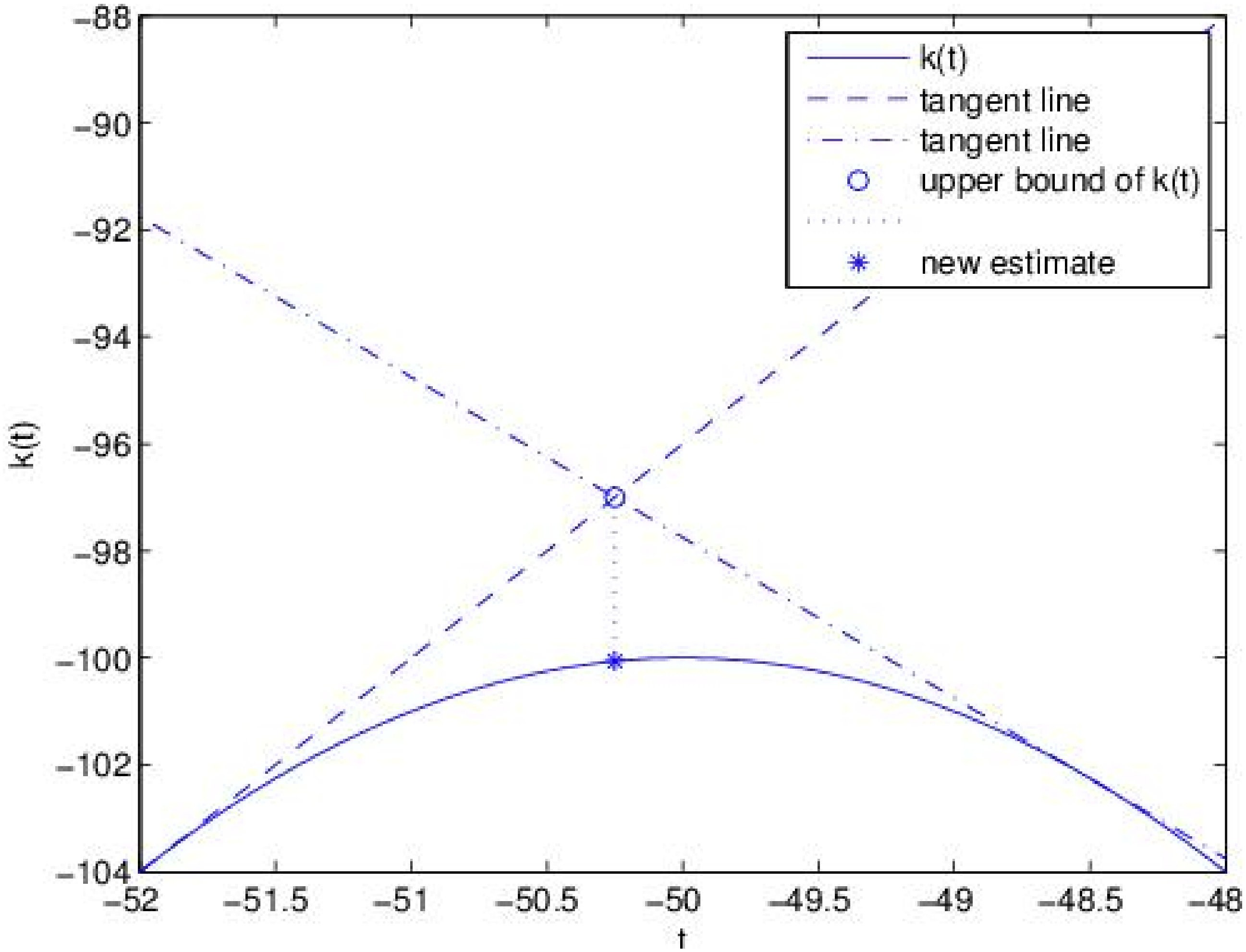
# Maximizing $k(t)$

- Triangle Interpolation.
- Vertical Cut.
- Inverse Linear Interpolation: consider

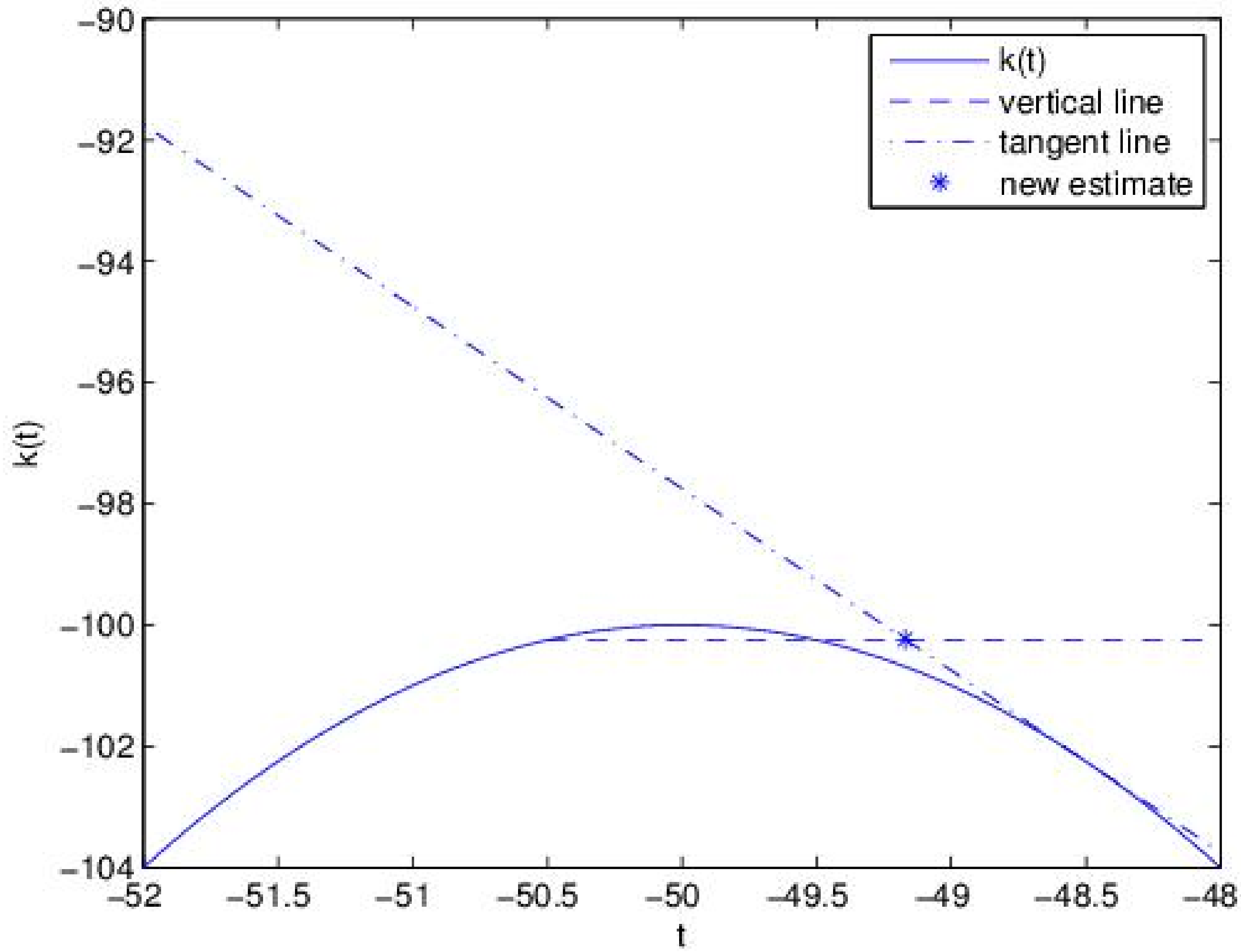
$$\Psi(t) := \sqrt{|s+1|} - \frac{1}{y_0(t)}.$$

Approximate the inverse of $\Psi$ by a linear function $t(\Psi) = a\Psi + b$, and set $t_+ = t(0)$.

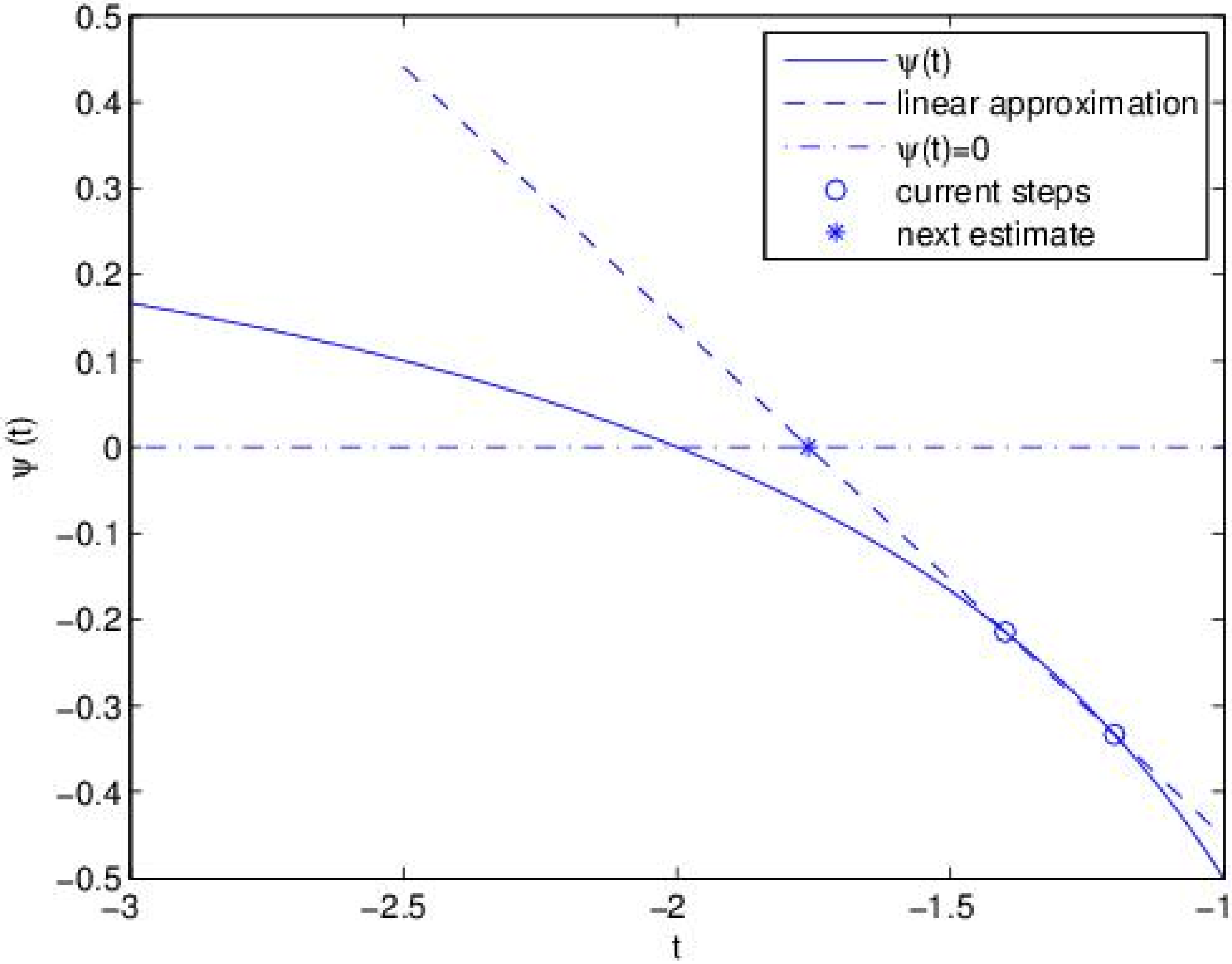
# Triangle Interpolation



# Vertical Cut



# Inverse Linear Interpolation



# Simulations

- Assume $B \succ 0$ and $b = 0$; consequently $s_- = 0 > -1$. One can show that

$$\overline{\lambda} - \sqrt{\frac{a^T B^{-1} a}{s}} \leq t^* \leq \overline{\lambda} + \sqrt{s a^T B^{-1} a}$$

# Simulations

- Assume $B \succ 0$ and $b = 0$; consequently $s_- = 0 > -1$. One can show that

$$\bar{\lambda} - \sqrt{\frac{a^T B^{-1} a}{s}} \le t^* \le \bar{\lambda} + \sqrt{s a^T B^{-1} a}$$

- For each $n = 2000$, $4000$, $6000$, $8000$, $10000$, generate $10$ "easy instances", with known optimal solution:

```
B = sprandsym(n,0.01,0.01,2);
A = sprandsym(n,0.01);
lambdabar = eigifp(A,B,1,opts);
lambda = lambdabar - 5 - 5*rand(1);
x = randn(n,1); a = (A - lambda*B)*x; s = x'*B*x;
```

- Use eigifp (Golub, Ye '02) to compute generalized eigenvalues.

- Termination:

$$|k'(t)| < 10^{-7}(s+1) \text{ and } \frac{|q_0(x) - k(t)|}{|q_0(x)| + 1} < 10^{-14} \text{ and } \frac{|q_1(x) - s|}{s+1} < 10^{-14}$$

- CPU: Intel i5-2430M 2.4 GHz; RAM: 8 GB; Matlab: 7.12.

Table 1: Performance of RW algorithm, averaged over $10$ instances.

| $n$ | time | iter | $\frac{\|q_1(x_{\text{final}})-s\|}{s+1}$ | $\frac{\|q_0(x_{\text{final}})-\inf_{q_1(x)=s} q_0(x)\|}{\|q_0(x_{\text{final}})\|+1}$ |
|---|---|---|---|---|
| 2000 | 1.7 | 8 | 2.89e-18 | 2.43e-16 |
| 4000 | 3.2 | 8 | 1.20e-18 | 1.10e-16 |
| 6000 | 5.2 | 8 | 1.50e-17 | 2.45e-16 |
| 8000 | 7.6 | 8 | 9.57e-17 | 2.23e-16 |
| 10000 | 11.1 | 8 | 1.55e-16 | 3.20e-16 |

# Conclusion & Future work

- GTRS can be analyzed similarly as TRS, under suitable assumptions.
- The RW algorithm can be extended to solve GTRS.
- More numerical tests...
- Estimating/bounding $t_-$ when it is finite.

Thanks for coming!