Generalized Trust Region Subproblem: Analysis and Algorithm

Ting Kei Pong
Combinatorics & Optimization, University of Waterloo
Waterloo

Optimization Days
May 2013
(Joint work with Henry Wolkowicz)

Outline

- Generalized trust region subproblem.
- Optimality conditions.
- Easy case and hard cases.
- (Extended) Rendl-Wolkowicz algorithm.
- 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. The dual problem is strictly feasible; i.e.,

$$\exists \lambda \text{ s.t. } A - \lambda B \succ 0.$$

- The following constraint qualification holds

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

- *Nondegeneracy* assumption:

$$\begin{bmatrix} a \\ b \end{bmatrix} \notin \mathcal{R}\left(\begin{bmatrix} A \\ B \end{bmatrix}\right)$$

Implications

- GTRS has an optimal solution. Moreover:

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

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

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

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

- $\exists$ invertible matrix S s.t. $A = S\mathrm{diag}(\alpha)S^T$ and $B = S\mathrm{diag}(\beta)S^T$.

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

Easy/Hard Cases for GTRS

For simplicity, we consider $s := u = l$, $B \succeq 0$ or indef.

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

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

Extended Rendl-Wolkowicz Algorithm

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

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

where

$$\begin{aligned} k_0(t) &:= \inf_{x^T Bx - 2b^T xy_0 + y_0^2 = s+1} x^T Ax - 2a^T xy_0 + ty_0^2 \\ &= \inf \left\{ y^T \begin{bmatrix} t & -a^T \\ -a & A \end{bmatrix} y : \; y^T \begin{bmatrix} 1 & -b^T \\ -b & B \end{bmatrix} y = s+1 \right\}. \end{aligned}$$

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

- 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.

Intuition behind $k(t)$

- Roughly speaking, $k(t) = |s+1|d^{-1}(t) - t$.
- Here, $d(\lambda)$ is the dual obj. for GTRS $\min_{x^T Bx - 2b^T x = -1} x^T Ax - 2a^T x$:

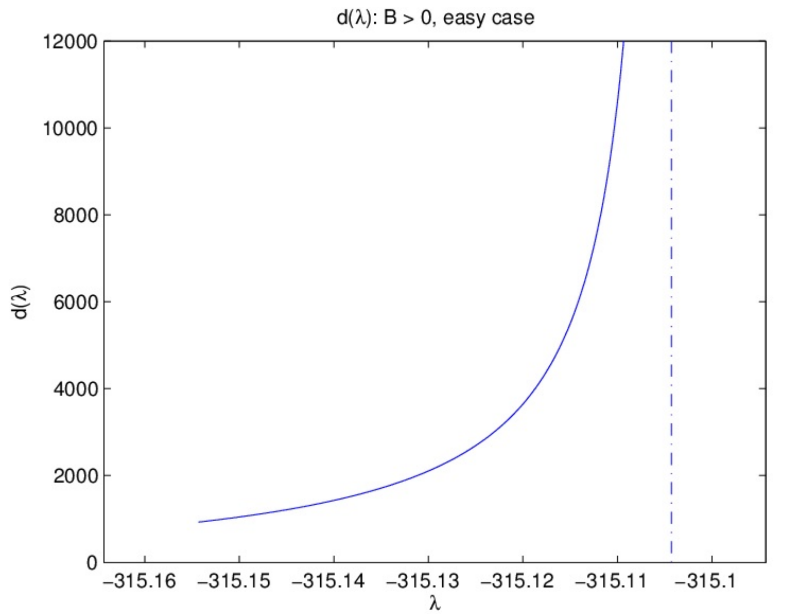
  $d(\lambda) = \lambda + (a - \lambda b)^T (A - \lambda B)^{-1}(a - \lambda b)$, defined for $A - \lambda B \succ 0$.

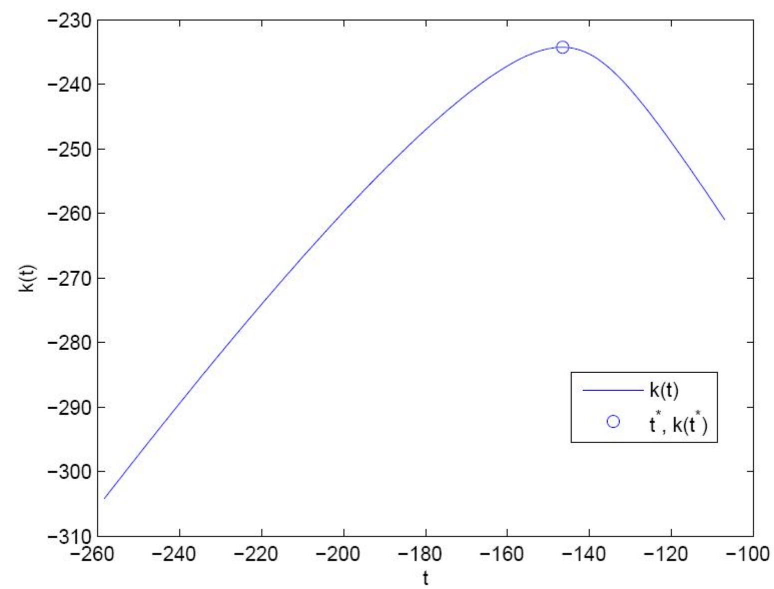
Intuition behind $k(t)$

- Roughly speaking, $k(t) = |s+1|d^{-1}(t) - t$.

- Here, $d(\lambda)$ is the dual obj. for GTRS $\min_{x^T Bx - 2b^T x = -1} x^T Ax - 2a^T x$:

  $d(\lambda) = \lambda + (a - \lambda b)^T (A - \lambda B)^{-1}(a - \lambda b)$, defined for $A - \lambda B \succ 0$.





$B \succ 0$, easy case, $\{x : \ x^T Bx - 2b^T x = -1\} = \emptyset$.

Properties of $k(t)$

Fact 4 (P, Wolkowicz '12): Suppose $s + 1 \neq 0$. Let

$$t_- := \inf_{A - \lambda B \succ 0} \lambda + (a - \lambda b)^T (A - \lambda B)^{-1} (a - \lambda b),$$
$$s_- := \inf x^T B x - 2b^T x.$$

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

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

Easy/Hard Case 1: Solution Recovery

Fact 5 (P, Wolkowicz '12): Suppose $s+1 \neq 0$ and $t^* = \operatorname{argmax} k(t)$.

Then $t^* > t_-$. The generalized eigenvalue $\lambda^* = d^{-1}(t^*) \in (\underline{\lambda}, \overline{\lambda})$ of

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

is simple; any generalized eigenvector $y(t^*) = \begin{bmatrix} y_0(t^*) \\ w(t^*) \end{bmatrix}$ satisfies $y_0(t^*) \neq 0$.

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

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

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 Ψ by a linear function $t(\Psi) = a\Psi + b$, and set $t_+ = t(0)$.

Simulations: $B \succ 0$, $b = 0$, indef. A, easy case

- Use eigifp (Golub, Ye '02) to compute generalized eigenvalues.
- For RW algorithm: $s_- = 0 > -1$; Also, we have (P, Wolkowicz '12)

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

  Use these end points to initialize RW algorithm.
- Compare with Newton + Armijo on minimizing the dual:

$$\min_{A - \lambda B \succ 0} -s\lambda + a^T (A - \lambda B)^{-1} a,$$

  initialized at $\lambda^0 = \overline{\lambda} - 1$.

Simulations: $B \succ 0$, $b = 0$, indef. A, easy case

For each $n = 10000, 15000, 20000$, generate 10 "easy instances":

```
B = sprandsym(n,1e-2,0.1,2); A = sprandsym(n,1e-2);
lambda = eigifp(A,B,1,opts) - 5 - 5*rand(1);
x = randn(n,1)/10; a = (A - lambda*B)*x; s = x'*B*x;
u = 1.2*s; l = 0.8*s;
```

Apply Newton to dual function with $s = u$.

Table 1: Indef. A and $B \succ 0$.

n	RW iter/cpu/fval/feas$_{\mathbf{eq}}$	Newton+Armijo iter/cpu/fval/feas$_{\mathbf{eq}}$
10000	6/4.92/-4.106880221e+4/4.1e-13	9/6.89/-4.106880222e+4/5.5e-07
15000	6/12.89/-9.910679264e+4/5.5e-13	9/17.56/-9.910679264e+4/1.0e-07
20000	6/19.26/-1.443904856e+5/1.6e-12	9/29.85/-1.443904856e+5/8.7e-07

Conclusion & Future work

- GTRS can be analyzed similarly as TRS, under suitable assumptions.
- The RW algorithm can be extended to solve GTRS.
- Open problem: estimating/bounding t_- when it is finite.

Thanks for coming! $\ddot\smile$

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

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

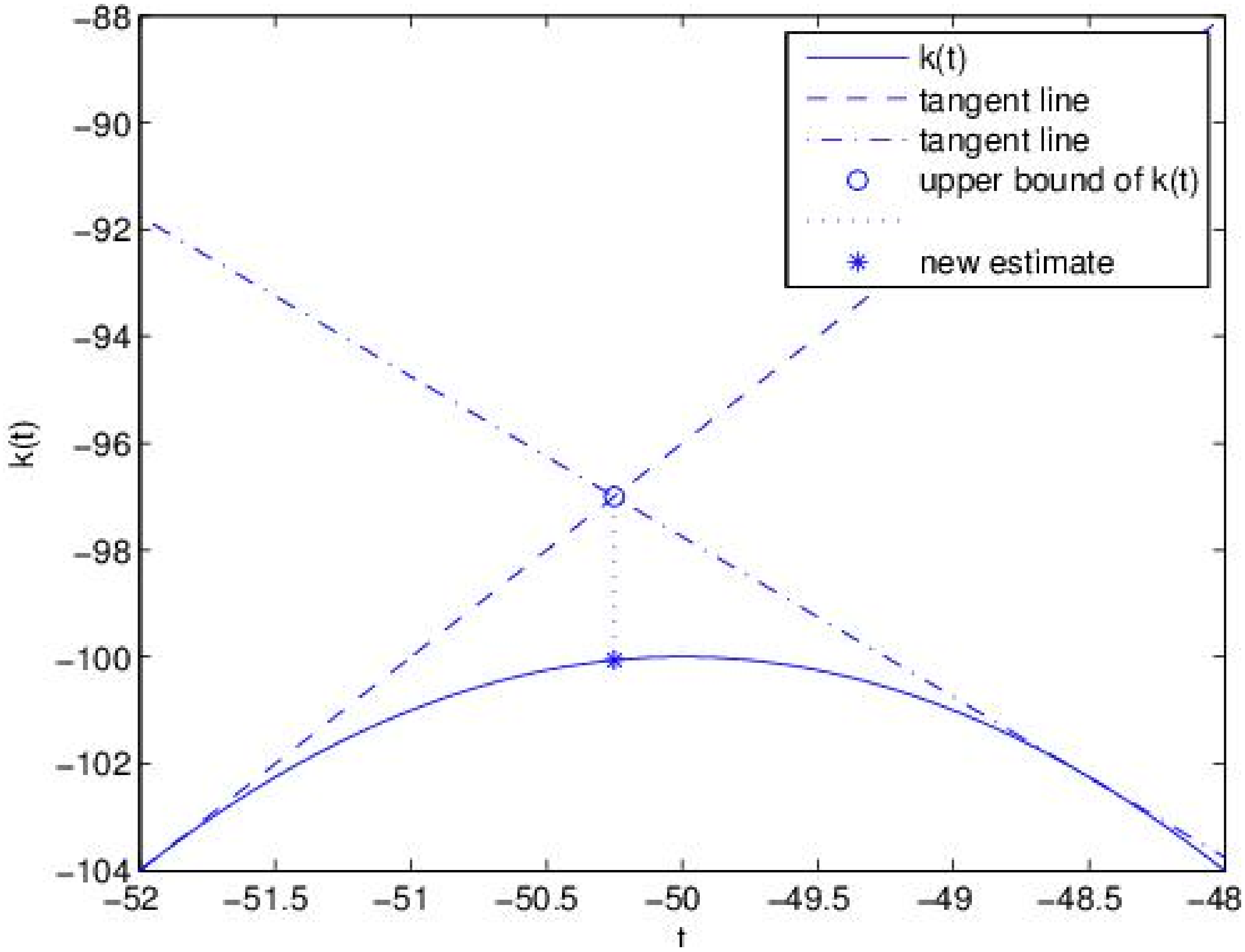
Similarly, if $\underline{\lambda}$ is finite, then

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

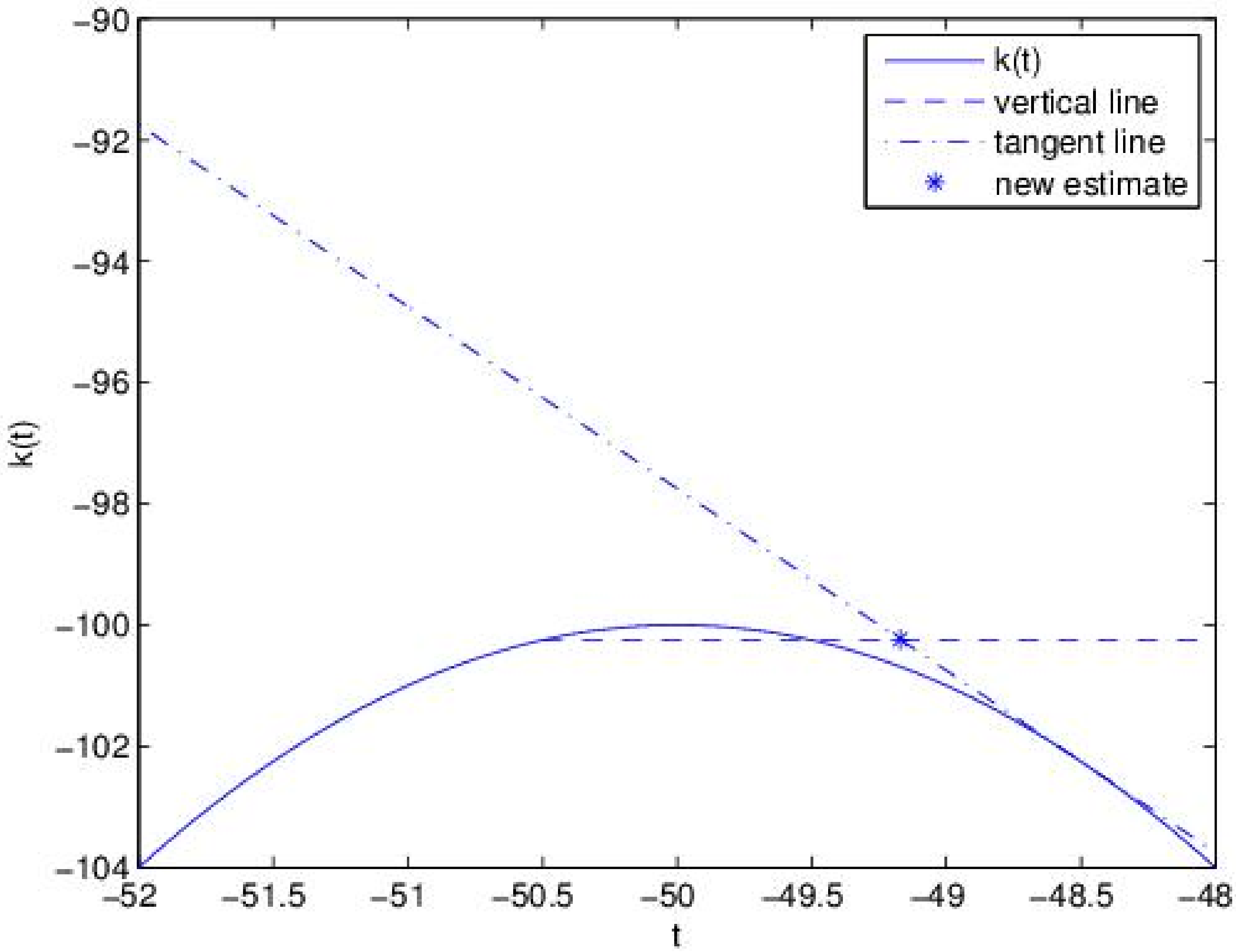
Since ψ is non-decreasing on $(\underline{\lambda}, \overline{\lambda})$,

- 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 $v \in \mathcal{N}(A - \underline{\lambda} B)$.

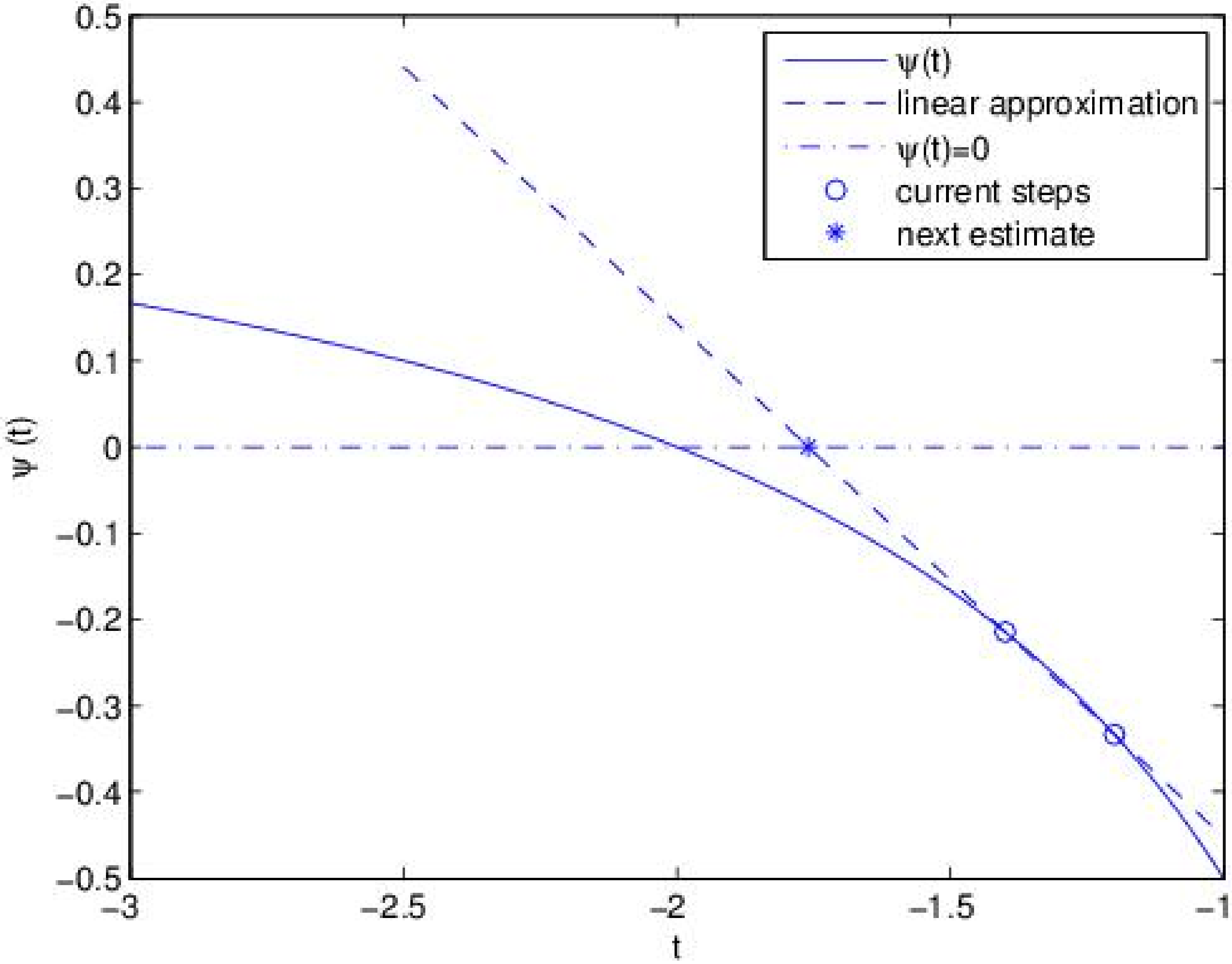
Triangle Interpolation



Vertical Cut



Inverse Linear Interpolation



Termination Criteria & Hardware Info.

- Termination for RW algorithm:

$$\max \left\{ \frac{|k'(t)|^2}{|s+1|^2}, \frac{|q_0(x) - k(t)|}{|q_0{}^{\text{best}}| + 1}, \frac{|q_1(x) - s|}{|s| + 1}, \frac{\|Ax - \lambda Bx - a\|^2}{(\|A\|_2 + \|a\| + 1)^2} \right\} < 10^{-13}$$

- Termination for Newton + Armijo on dual function, termination:

$$\max \left\{ \frac{|q_0(x) - d(\lambda)|}{|q_0(x)| + 1}, \frac{|q_1(x) - s|}{|s| + 1}, \frac{\|Ax - \lambda Bx - a\|^2}{(\|A\|_2 + \|a\| + 1)^2} \right\} < 10^{-12}.$$

- Two 2.4 GHz quad-core Intel E5620 Xeon 64-bit CPUs, 48 GB RAM, Matlab 7.14 (R2012a).