Puiseux-based Extrapolation for Large-Scale Degenerate Quadratic Programming

Nick Gould
STFC Rutherford Appleton Laboratory
with
Dominique Orban (Montreal) & **Daniel Robinson** (Northwestern)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} x^T H x + g^T x \quad \text{subject to} \quad Ax = b \quad \text{and} \quad x \geq 0$$

24th Biennial Conference on Numerical Analysis
University of Strathclyde, Scotland, 29th June 2011

Problem

QP: $\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} x^T H x + g^T x \quad \text{subject to} \quad Ax = b \;\; \& \;\; x \geq 0$

- assume that A has full row rank & $H \succeq 0$
- aim to (approximately) satisfy **criticality conditions**

$$\begin{aligned} Ax_* = b \;\&\; x_* &\geq 0 && \text{(primal feasibility)} \\ g + Hx_* - A^T y_* - z_* = 0 \;\&\; z_* &\geq 0 && \text{(dual feasibility)} \\ x_* \cdot z_* &= 0 && \text{(complementary slackness)} \end{aligned}$$

or to deduce that the problem is infeasible

- problem **non degenerate** $\iff$ $\exists$ solution s.t. $\max(x_{*,i}, z_{*,i}) > 0$ $\forall i$ ($\iff$ a strictly complementary solution)
- problem **degenerate** $\iff$ not non-degenerate!
- aim is to find highly-accurate solutions even when QP is degenerate

Generic path following strategy

Given $(x_0, z_0) > 0$ and y_0, trace the (infeasible) **trajectory**

$$v(\mu) = (x(\mu), y(\mu), z(\mu))$$

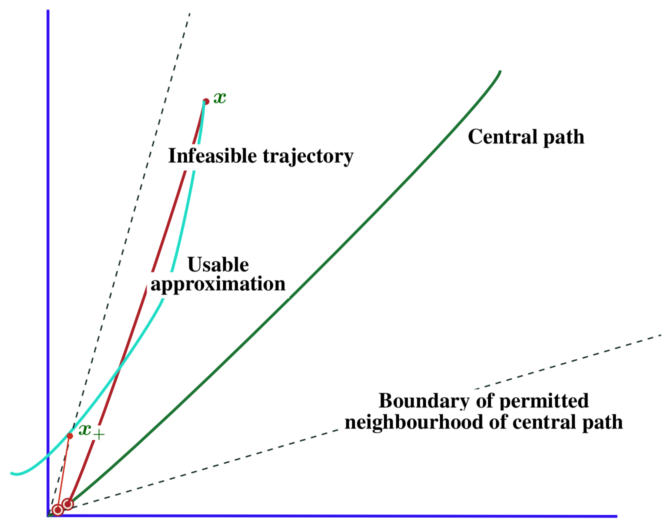
where

$$\begin{aligned} Ax(\mu) - b &= \mu[Ax_0 - b] \\ g + Hx(\mu) - A^T y(\mu) - z(\mu) &= \mu[g + Hx_0 - A^T y_0 - z_0] \\ x(\mu) \cdot z(\mu) &= c(\mu) \end{aligned}$$

with $c(1) = x_0 \cdot z_0$ and $c(0) = c_{\text{target}} \approx 0$ as μ decreases from 1 to 0

- usually achieve this using a suitably safeguarded Newton (i.e., Taylor series-based) iteration
- adjust c_{target} for convergence

A generic algorithm in pictures



Example 1 - non-degenerate QP

$$\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^2 \quad \text{subject to} \quad x \geq 2$$

trajectory ($c(\mu) = \mu$): $x(\mu) = 1 + \sqrt{1+\mu}$ $\longleftarrow$ analytic for $\mu \geq 0$

Iter	p-feas	d-feas	com-slk	obj	step	mu	arc
0	0.0E+00	1.0E+00	2.0E+00	4.5E+00	-	2.0E-02	-
1r	0.0E+00	0.0E+00	1.3E-02	2.0E+00	1.0E+00	1.3E-04	1TZh
2r	0.0E+00	4.4E-16	1.8E-04	2.0E+00	1.0E+00	1.8E-06	1TZh
3r	0.0E+00	0.0E+00	1.8E-06	2.0E+00	1.0E+00	2.4E-09	1TZh
4r	0.0E+00	4.4E-16	2.4E-09	2.0E+00	1.0E+00	1.2E-13	1TZh
5r	0.0E+00	4.4E-16	1.2E-13	2.0E+00	1.0E+00	3.9E-20	1TZh

Example 2 - degenerate QP

$$\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^2 \quad \text{subject to} \quad x \geq 0$$

trajectory: $x(\mu) = \sqrt{\mu}$ $\longleftarrow$ not analytic at 0

Iter	p-feas	d-feas	com-slk	obj	step	mu	arc
0	0.0E+00	1.0E+00	2.0E+00	5.0E-01	-	2.0E-02	-
1r	0.0E+00	0.0E+00	4.5E-01	2.3E-01	1.0E+00	4.5E-03	1TZh
2r	0.0E+00	0.0E+00	1.2E-01	5.8E-02	1.0E+00	1.2E-03	1TZh
3r	0.0E+00	0.0E+00	2.9E-02	1.5E-02	1.0E+00	2.9E-04	1TZh
4r	0.0E+00	0.0E+00	7.5E-03	3.8E-03	1.0E+00	7.5E-05	1TZh
5r	0.0E+00	0.0E+00	1.9E-03	9.6E-04	1.0E+00	1.9E-05	1TZh
6r	0.0E+00	3.5E-18	4.9E-04	2.4E-04	1.0E+00	4.9E-06	1TZh
...	....	....	....	....	....	....	....
18r	0.0E+00	0.0E+00	3.1E-11	1.6E-11	1.0E+00	1.7E-16	1TZh
19r	0.0E+00	0.0E+00	7.8E-12	3.9E-12	1.0E+00	2.2E-17	1TZh
20r	0.0E+00	6.4E-22	1.9E-12	9.7E-13	1.0E+00	2.7E-18	1TZh
21r	0.0E+00	3.2E-22	4.8E-13	2.4E-13	1.0E+00	3.4E-19	1TZh

Example 2 again

$$\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^2 \quad \text{subject to} \quad x \geq 0$$

re-parameterize trajectory: $\mu = \rho^2 \longrightarrow x(\rho) = \rho$ $\longleftarrow$ analytic

Iter	p-feas	d-feas	com-slk	obj	step	mu	arc
0	0.0E+00	1.0E+00	2.0E+00	5.0E-01	-	2.0E-02	-
1r	0.0E+00	1.1E-16	5.6E-01	2.8E-01	1.0E+00	5.6E-03	1PZh
2r	0.0E+00	1.1E-16	5.6E-05	2.8E-05	1.0E+00	4.2E-07	1PZh
3r	0.0E+00	1.1E-16	3.1E-09	1.5E-09	1.0E+00	1.7E-13	1PZh
4r	0.0E+00	1.1E-16	9.6E-18	4.8E-18	1.0E+00	3.0E-26	1PZh

What is the fundamental difference?

- use a **Puiseux** rather than Taylor approximation to the trajectory

Non-degenerate QP

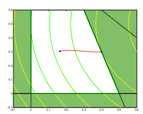
For simplicity

- suppose v_0 is a strictly feasible primal-dual interior point
- consider the (weighted) **central path** $v(\mu)$ as $\mu \rightarrow 0_+$:

$$\begin{aligned} Ax(\mu) - b &= 0 \\ g + Hx(\mu) - A^T y(\mu) - z(\mu) &= 0 \\ x(\mu) \cdot z(\mu) &= \mu x_0 \cdot z_0 \end{aligned}$$

Then $v(\mu)$ is analytic at 0 whenever QP is non-degenerate

(Fiacco & McCormick)



$\Longrightarrow$ Taylor series-based methods work

- higher-order Taylor approximations are possible by differentiating central path equations
- always works for **linear** programming

Degenerate QP

For simplicity

- suppose v_0 is a strictly feasible primal-dual interior point
- consider the re-parameterized central path $v(\rho)$ as $\rho \rightarrow 0_+$:

$$\begin{aligned} Ax(\rho) - b &= 0 \\ g + Hx(\rho) - A^T y(\rho) - z(\rho) &= 0 \\ x(\rho) \cdot z(\rho) &= \rho^2 x_0 \cdot z_0 \end{aligned}$$

Then $v(\rho)$ has an analytic extension at 0 even if QP is degenerate

(Stoer, Wechs & Mizuno, 1998)

$\Longrightarrow$ Taylor series-based methods work for this parameterization

- higher-order Taylor approximations are possible by differentiating re-parameterized central path equations
- returning to the original μ parametrization leads to a **Puiseux** series

Puiseux series

Taylor series representation of the re-parameterized central path

$$v(\rho) = \sum_{i \geq 0} v^{[i]} \frac{(\rho - \rho_k)^i}{i!}$$

about (ρ_k, v_k) where $\mu_k = \rho_k^2$ becomes the **Puiseux** series

$$v(\mu) = \sum_{i \geq 0} v^{[i]} \frac{(\sqrt{\mu} - \sqrt{\mu_k})^i}{i!}$$

Coefficients $v^{[i]}$ found by solving a sequence of **primal-dual** systems

$$\begin{pmatrix} H & -A^T & -I \\ A & 0 & 0 \\ Z_k & 0 & X_k \end{pmatrix} \begin{pmatrix} x^{[i]} \\ y^{[i]} \\ z^{[i]} \end{pmatrix} = r_i(v^{[0]}, \ldots, v^{[i-1]})$$

for easily-determined rhs $r_i(v^{[0]}, \ldots, v^{[i-1]})$, where $v^{[0]} = v_k$

- odd-order coefficients $\rightarrow 0$ in non-degenerate case

Local trajectory

Locally use the trajectory $v_k(\mu)$, where

$$\begin{aligned} Ax_k(\mu) - b &= (\mu/\mu_k)[Ax_k - b] \\ g + Hx_k(\mu) - A^T y_k(\mu) - z_k(\mu) &= (\mu/\mu_k)[g + Hx_k - A^T y_k - z_k] \\ x_k(\mu) \cdot z_k(\mu) &= c_k(\mu) \end{aligned}$$

for which $c_k(\mu_k) = x_k \cdot z_k$, as μ decreases from μ_k to 0

- different c_k give different trajectories $\Longrightarrow$ choice important

Choice of the complementarity function I

Choice of c_k for which $c_k(\mu_k) = x_k \cdot z_k$ and

$$x_k(\mu) \cdot z_k(\mu) = c_k(\mu)$$

leads to different trajectories:

1. **linear interpolation to the central path**

$$c_k(\mu) = \frac{\mu}{\mu_k} x_k \cdot z_k + \left(1 - \frac{\mu}{\mu_k}\right) \sigma_k \frac{x_k^T z_k}{n} e$$

with $0 \leq \sigma_{\min} \leq \sigma_k \leq \sigma_{\max} < 1$ (Zhang, 1994)

- interpolates $x_k \cdot z_k$ and $\sigma_k (x_k^T z_k / n) e$
- Taylor coefficients may diverge for degenerate QP
- may prefer Puiseux $\mu = \rho^2$ alternative

Choice of the complementarity function II

2. **quadratic interpolation to the solution**

$$c_k(\mu) = \frac{\mu}{\mu_k} x_k \cdot z_k + \mu \left(1 - \frac{\mu}{\mu_k}\right) \left(\frac{x_k^T z_k}{n} e - x_k \cdot z_k\right)$$

(Zhao & Sun, 1999, Potra & Stoer, 2009)

- interpolates $x_k \cdot z_k$ and $\mathbf{0}$ but crucially ensures bounded $c'(0)$ $\Longrightarrow$ bounded leading Taylor coefficient
- may prefer Puiseux $\mu = \rho^2$ alternative (Potra & Stoer)
- simpler Puiseux variant

$$\frac{\rho^2}{\mu_k} x_k \cdot z_k + \frac{\rho^2}{\mu_k} (\sqrt{\mu_k} - \rho) \left(\frac{x_k^T z_k}{n} e - x_k \cdot z_k\right)$$

also possible (Zhao & Sun)

Basic algorithm

- approximate $v_k(\mu)$ by an ℓ th order Taylor or Puiseux series, $v_k^{(\ell)}(\mu)$
- for appropriate κ_c and κ_f, define the the $\mathcal{N}_\infty^-$**-neighbourhood**:

$$[x \cdot z]_i \geq \kappa_c x^T z \quad \text{for all} \quad i$$

and

$$x^T z \geq \kappa_f \left\| \begin{pmatrix} Ax - b \\ g + Hx - A^T y - z \end{pmatrix} \right\|$$

- find the smallest $\mu_k^{\min} \in [0, \mu_k]$ for which $v_k^{(\ell)}(\mu) \in \mathcal{N}_\infty^-$ for all $\mu \in [\mu_k^{\min}, \mu_k]$ (Zhang, Wright, ...)
- find

$$\mu_{k+1} \approx \arg \min_{\mu \in [\mu_k^{\min}, \mu_k]} x_k^{(\ell)T}(\mu)\, z_k^{(\ell)}(\mu)$$

and set $v_{k+1} = v_k^{(\ell)}(\mu_{k+1})$

More sophisticated algorithm

- apply the basic algorithm for a variety of Taylor and/or Puiseux series and complementarity functions
  - coefficients for lower-order series automatically available from highest-order one
- pick the one that gives the smallest complementarity
- to ensure convergence, include 1st-order Taylor-Zhang
  - polynomial algorithm (Zhang, 1994, Billups & Ferris, 1996)
- to ensure fast convergence, include ℓ th-order Puiseux-Zhao-Sun
  - ultimately Q-order $(\ell + 1)/2$ (Zhao & Sun, 1999, Potra & Stoer, 2009)
  - improved polynomial bound (Potra & Stoer)
  - for non-degenerate problems ℓ th-order Taylor-Zhao-Sun gives ultimately Q-order $\ell + 1$

Credit where credit is due

- many of these ideas originated in linear complementarity during the 1990s and 2000s
  - usually first for monotone LCP
  - then generalised for sufficient LCP
- large number of papers, without exception theoretical and with no practical evaluation
- key players include Kojima, Mizuno, Noma, (Y&Z) Zhang, Billups, Ferris, Wright, Stoer, Wechs, Sturm, Liu, Potra, Zhao, Sun, ...

An implementation: CQP

Implemented as module CQP as part of GALAHAD

- fortran 2003 with many options
  - general $x^L \leq x \leq x^U$ & $c^L \leq Ax \leq c^U$ allowed
    - infinite and/or duplicated bounds permitted $\Longrightarrow$ free variables, one-sided constraints, equalities, etc
  - choice of pre-scaling schemes
  - dependent constraint removal & pre-solve
  - choice of linear solver
  - choice of complementarity function to define trajectory
  - Taylor or Puiseux series of specified order
  - can try lower orders as well
  - optimal active-set indicators obtained
  - crossover (coming)
- available without cost for non-incorporational use

galahad.rl.ac.uk

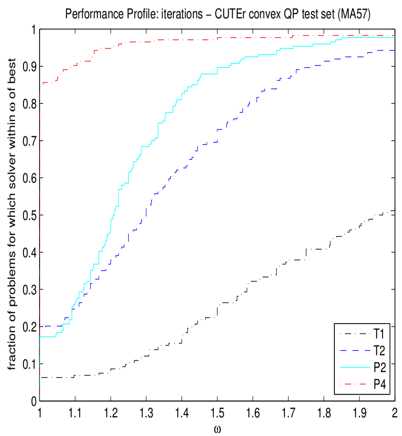
Dominant costs

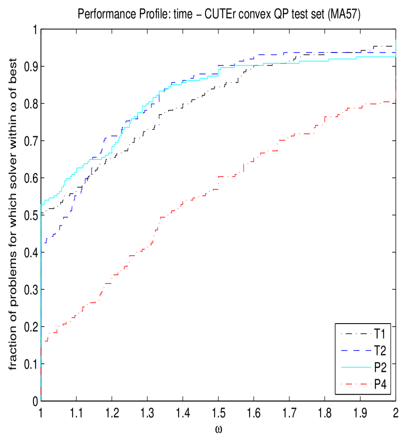
- build Taylor/Puiseux approximations
  - factorize primal-dual matrix

$$\begin{pmatrix} H & -A^T & -I \\ A & 0 & 0 \\ Z_k & 0 & X_k \end{pmatrix}$$

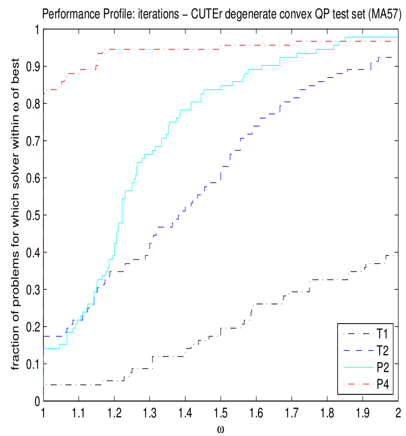
    - uses GALAHAD's linear equation über-solver SLS with access to MA57, out-of-core MA77, and parallel MA86/87/97/PARDISO & WSMP, etc
  - solve ℓ systems with this to obtain coefficients
    - ratio of solves/factorize poor on multicore CPUs ☹
- find the maximum stepsize in the $\mathcal{N}_\infty^-$-neighbourhood
  - find appropriate roots of $2n+1$ univariate real polynomials each of degree 2ℓ in **parallel**
    - use efficient Sturm-sequence iteration using GALAHAD's ROOTS
- matrix-vector products with H, A and A^T

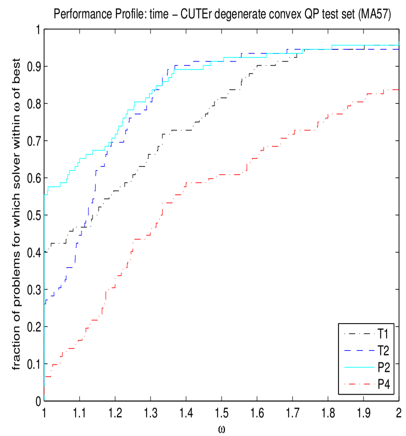
Performance profiles





Performance profiles — degenerate problems





Summary

- highly accurate solution of degenerate QP is generally not possible by standard higher order methods
- degeneracy may be overcome by using a “square-root” Puiseux expansion based on an analytic re-parameterization
- polynomial and superlinear convergence is possible in all cases
- extends to classes of LCP
- higher order Puiseux expansions improve the number of factorizations required, but the time savings may be outweighed by the number of linear (forward & back) solves required
- GALAHAD solver `CQP` available
- also used as a heuristic in the nonconvex QP solver `QPC`