A Polynomial Oriented IPM Framework for Infeasible Iterates

F.R. Villas-Bôas, A.R.L. Oliveira,
C. Perin and L.R. Santos

fernandovillasboas@gmail.com

University of Campinas

keywords: IPM, Predictive polynomials,
Infeasible Central path and neighborhood

IPM Context

- Linear Programming (LP)
- Path-Following
- Using Merit Function
- Combining various directions
- Theory and Implementation

Key point: Framework Infeasible IPM

- Key tool: Polynomials
- Generalizes and integrates central path, neighborhoods, merit function
- Takes data size into account
- Computational results & proof-of-concept (theory equal to implementation)

Highlights:
Present IPM (simplifying...)

- Different directions combined to generate next iterate
- Full step correction may be excessive
- Good iterates near the Central Path
- Heuristics forcing iterates into neighborhood
- Merit function evaluates iterates

What we want (and did)

- Merit function to *guide* rather than evaluate
- Neighborhood as a set of constraints
- Combination of search directions as an optimization problem
- Practical implementation as proof-of-concept

KKT & Central Path

KKT :

$$\begin{cases} Ax - b = 0 \\ A^T y + z - c = 0 \\ xz = 0 \\ x, z \geq 0 \end{cases}$$

$\mathcal{C} = (x, y, z)$:

$$\begin{cases} Ax - b = 0 \\ A^T y + z - c = 0 \\ xz = \mu = \dfrac{x^T z}{n} \\ x, z \geq 0 \end{cases}$$

$$x, z, c \in \mathbb{R}^n, \quad y, b \in \mathbb{R}^m, \quad A \in \mathbb{R}^{m \times n}$$

Proposal: Scale original KKT (sign and magnitude)

- Sign based on the initial point
- Magnitude based on problem's data

$$h = \xi \varrho$$

$$\begin{cases} h_P^{-1}(Ax - b) = 0 \\ h_D^{-1}(A^T y + z) = 0 \\ xz = 0 \end{cases}$$

KKT scale

– Sign: $$\xi : \begin{cases} \xi_i^P = \begin{cases} +1, & (Ax_0 - b)_i \geq 0 \\ -1, & (Ax_0 - b)_i < 0 \end{cases} \\ \xi_i^D = \begin{cases} +1, & (A^T y_0 + z_0 - c)_i \geq 0 \\ -1, & (A^T y_0 + z_0 - c)_i < 0 \end{cases} \end{cases}$$

– Magnitude: $$\varrho : \begin{cases} \varrho_i^P = \max\left\{ \max_j \{|A_{i,j}|\}, |b_i| \right\} \\ \varrho_i^D = \max\left\{ \max_j \{|A_{j,i}|\}, |c_i| \right\} \end{cases}$$

Rationale of scaling

- Residues become
  - Nonnegative
  - Comparable in size

Generalizing Central Path: Scaled approach

KKT :

$$\begin{cases} Ax - b = 0 \\ A^T y + z - c = 0 \\ xz = 0 \\ x, z \geq 0 \end{cases}$$

$\mathcal{C} = (x, y, z)$:

$$\begin{cases} h_P^{-1}(Ax - b) = \mu \\ h_D^{-1}(A^T y + z - c) = \mu \\ xz = \mu \\ x, z \geq 0 \end{cases}$$

Search directions

– Solve for $(\delta_x, \delta_y, \delta_z)$: $\begin{cases} A\delta_x = 0 \\ A^T\delta_y + \delta_z = 0 \\ x\delta_z + z\delta_x = -xz \end{cases}$

– Solve for $(\Delta_x^\mu, \Delta_y^\mu, \Delta_z^\mu)$: $\begin{cases} A\Delta_x^\mu = h_P \\ A^T\Delta_y^\mu + \Delta_z^\mu = h_D \\ x\Delta_z^\mu + z\Delta_x^\mu = e \end{cases}$

– Solve for $(\Delta_x^\sigma, \Delta_y^\sigma, \Delta_z^\sigma)$: $\begin{cases} A\Delta_x^\sigma = 0 \\ A^T\Delta_y^\sigma + \Delta_z^\sigma = 0 \\ x\Delta_z^\sigma + z\Delta_x^\sigma = -\delta_x\delta_z \end{cases}$

Rationale of directions

$$\begin{bmatrix} \hat{x} \\ y \\ \hat{z} \end{bmatrix} = \begin{bmatrix} x + \alpha(\delta_x + \mu\Delta_x^\mu + \sigma\Delta_x^\sigma) \\ y + \alpha(\delta_y + \mu\Delta_y^\mu + \sigma\Delta_y^\sigma) \\ z + \alpha(\delta_z + \mu\Delta_z^\mu + \sigma\Delta_z^\sigma) \end{bmatrix}$$

solves the linear approximation

$$\begin{cases} h_P^{-1}(A\hat{x} - b) = \mu \\ h_D^{-1}(A^T y + \hat{z}) = \mu \\ x\Delta_z + z\Delta_x = \mu - \sigma\delta_x\delta_z \end{cases}$$

when $\alpha = 1$

Framework

- Vector of KKT residues : $\rho = \begin{bmatrix} h_P^{-1}(Ax - b) \\ h_D^{-1}(A^T y + z - c) \\ xz \end{bmatrix}$
- Average of residues (merit): $\varphi = \dfrac{1}{2n+m} \sum_{i=1}^{2n+m} \rho_i$
- Restrictions on residue dispersion around average (neighborhood) : $\dfrac{1}{\gamma}\varphi \geq \rho_i \geq \gamma \varphi$

Consequences 1

– Next iterate $\begin{bmatrix} \hat{x} \\ y \\ \hat{z} \end{bmatrix} = \begin{bmatrix} x + \alpha(\delta_x + \mu\Delta_x^\mu + \sigma\Delta_x^\sigma) \\ y + \alpha(\delta_y + \mu\Delta_y^\mu + \sigma\Delta_y^\sigma) \\ z + \alpha(\delta_z + \mu\Delta_z^\mu + \sigma\Delta_z^\sigma) \end{bmatrix}$

– Next vector of KKT residues

$$\hat{\rho}(\alpha,\mu,\sigma) = \begin{bmatrix} h_P^{-1}(A\hat{x} - b) \\ h_D^{-1}(A^T y + \hat{z} - c) \\ \hat{x}\hat{z} \end{bmatrix}$$

are both vector-valued polynomials in (α,μ,σ)

Consequences 2 (after replacements)

- Next residue becomes a vector-valued polynomial of (α,μ,σ)
- Merit (average of residues) becomes a real-valued polynomial of (α,μ,σ)
- Neighborhood becomes a set of $O(n)$ real polynomial restrictions
- Combination of directions can be posed as an optimization problem

Algebraic expression: Next vector of residues

$$\hat{\rho}(\alpha,\mu,\sigma) = \begin{bmatrix} (1-\alpha)\rho_P + \alpha\mu \\ (1-\alpha)\rho_D + \alpha\mu \\ (1-\alpha)\rho_C + \alpha\mu - \sigma\delta_x\delta_z + \alpha^2\Lambda(\mu,\sigma) \end{bmatrix}$$

where

$$\Lambda(\mu,\sigma) = \mu^2\Delta_x^\mu\Delta_z^\mu + \sigma^2\Delta_x^\sigma\Delta_z^\sigma + \mu(\Delta_x^\mu\delta_z + \Delta_z^\mu\delta_x) + $$
$$+ \mu\sigma(\Delta_x^\sigma\Delta_z^\mu + \Delta_x^\mu\Delta_z^\sigma) + \sigma(\Delta_x^\sigma\delta_z + \Delta_z^\sigma\delta_x) + \delta_x\delta_z$$

Predictive polynomials (total degree 4)

– Next nerit = Next average of residues

$$\hat{\varphi}(\alpha,\mu,\sigma) = \sum_{i=0}^{2}\sum_{j=0}^{2}\sum_{k=0}^{2} a_{i,j,k}\alpha^{i}\mu^{j}\sigma^{k}$$

– Inequality restrictions on the next iterate

$$\frac{1}{\gamma}\hat{\varphi}(\alpha,\mu,\sigma) \geq \hat{\rho}_i(\alpha,\mu,\sigma) \geq \gamma\hat{\varphi}(\alpha,\mu,\sigma)$$

Global Optimization Problem

$$\min_{\alpha,\mu,\sigma} \hat{\varphi}(\alpha, \mu, \sigma)$$

$$\frac{1}{\gamma}\hat{\varphi}(\alpha, \mu, \sigma) \geq \hat{\rho}_i(\alpha, \mu, \sigma) \geq \gamma\hat{\varphi}(\alpha, \mu, \sigma)$$

$$i = 1, \ldots, 2n + m$$

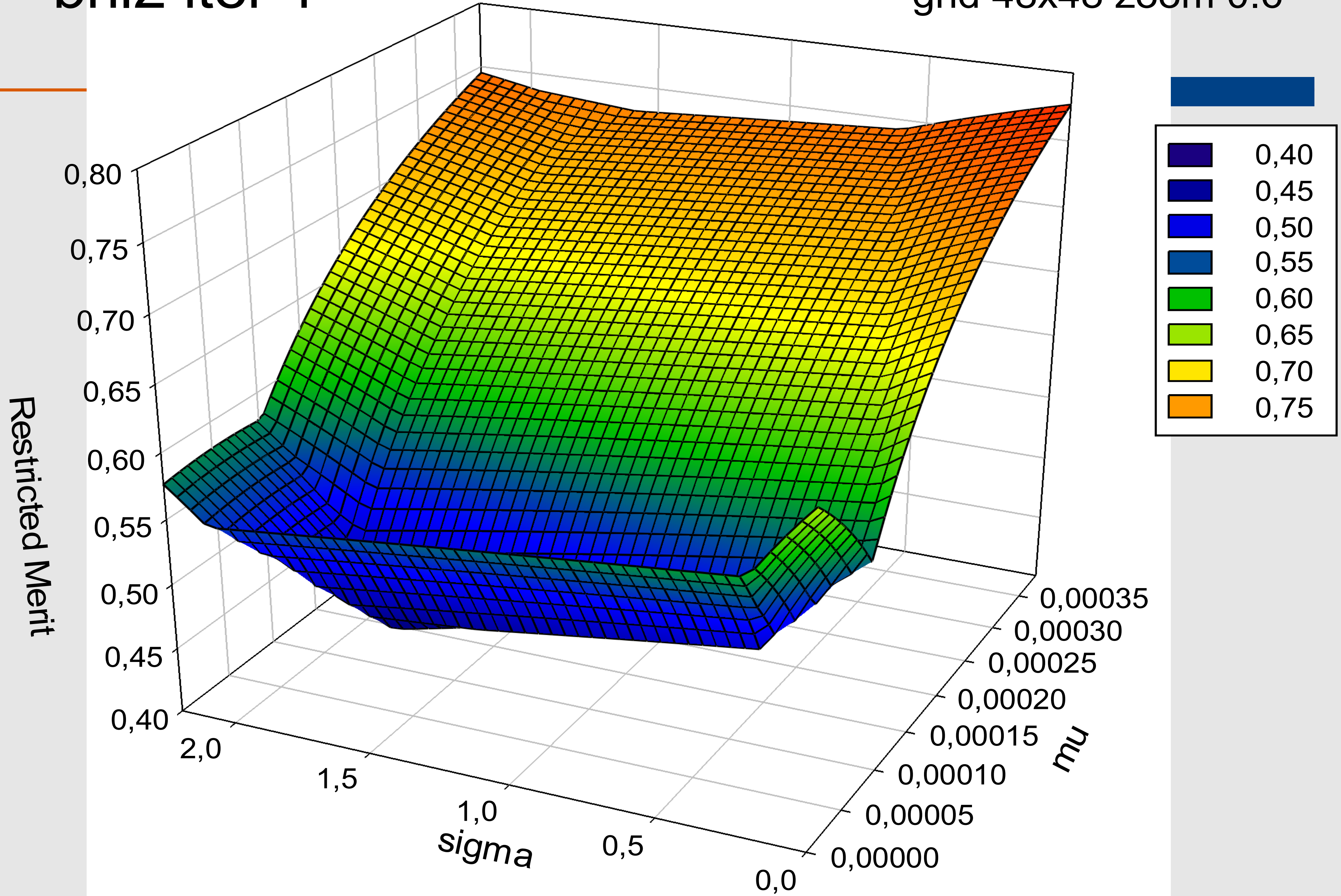
How we solve

- Approach 1 (implemented):
  approximate solution using cubic splines

- Approach 2 (under development):
  Split (μ,σ) domain into few regions (very few)
  where only one constraint is active and optimal
  can be found analytically

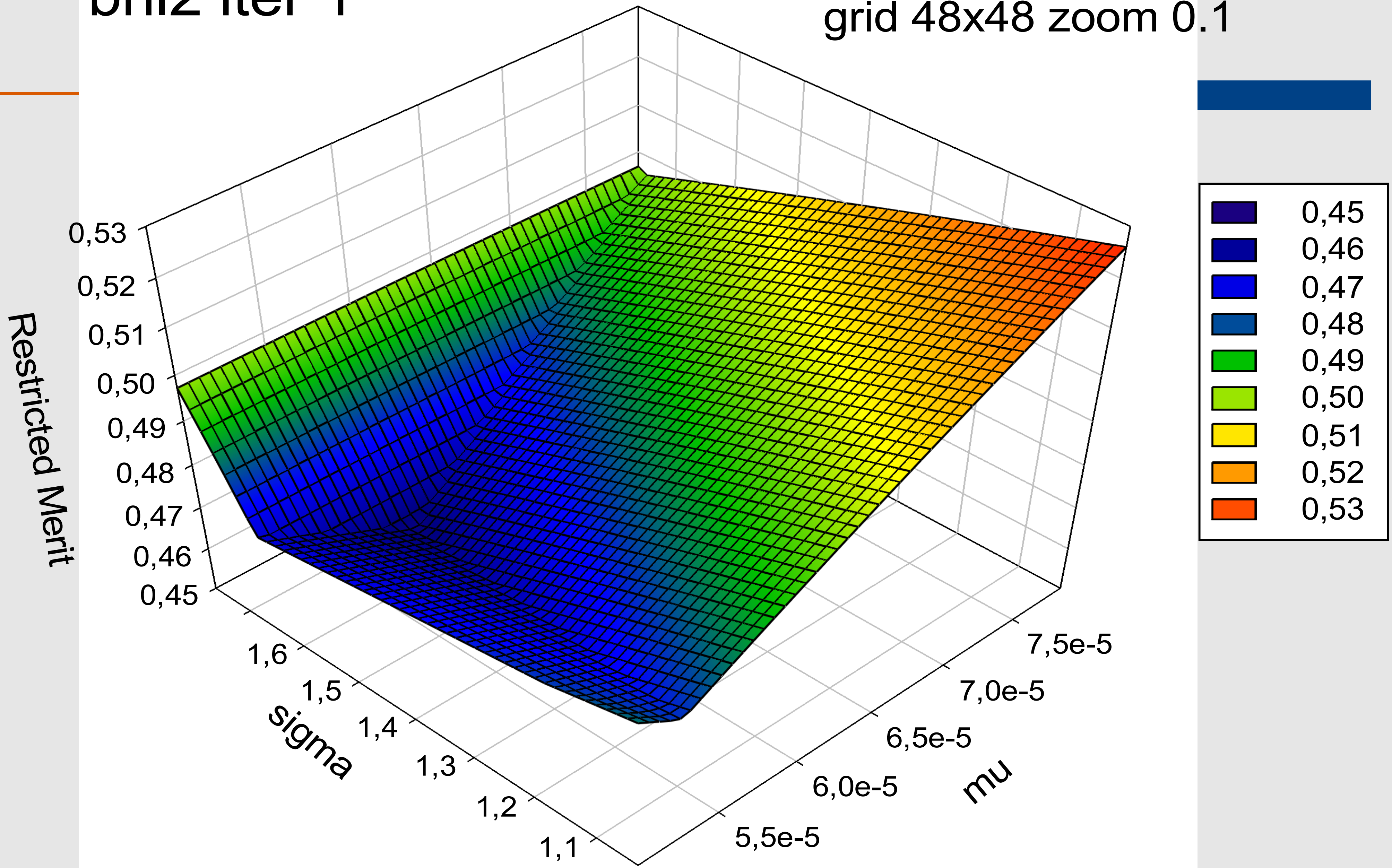
bnl2 iter 1

grid 48x48 zoom 0.6



bnl2 iter 1

grid 48x48 zoom 0.1



Problem Set: Netlib 108

25fv47
80bau3b
adlittle
afiro
agg
agg2
agg3
bandm
beaconfd
blend
bnl1
bnl2
boeing1
boeing2
bore3d
brandy
capri
cycle
czprob
d2q06c
d6cube
degen2
degen3
dfl001
e226
etamacro
fffff800
finnis
fit1d
fit1p
fit2d
fit2p
forplan
ganges
gfrd-pnc
greenbea
greenbeb
grow7
grow15
grow22
israel
kb2
lotfi
maros-r7
maros
modszk1
nesm
perold
pilot-ja
pilot-we
pilot
pilot4
pilot87
pilotnov
recipe
sc50a
sc50b
sc105
sc205
scagr7
scagr25
scfxm1
scfxm2
scfxm3
scorpion
scrs8
scsd1
scsd6
scsd8
sctap1
sctap2
sctap3
seba
share1b
share2b
shell
ship04l
ship04s
ship08l
ship08s
ship12l
ship12s
sierra
stair
standata
standgub
standmps
stocfor1
stocfor2
stocfor3
truss
tuff
vtp-base
wood1p
woodw
cre-a
cre-b
cre-c
cre-d
ken-07
ken-11
osa-07
osa-14
osa-30
pds-02
pds-06
pds-10
qap8

Comparison: PCx-C *versus* 998r-97

	PCx-C	998r-97	%
Time (s)	115.17	150.59	31%
Problems Solved	105	103	2%

A strange remark

- In this implementation 998r-97,
  for all problems in the problem set,
  and in all iterations,
  the optimal combination of directions had

$$\mu = 0$$

Conclusions

- Framework proposed
  - Predictive polynomials
  - Scaled KKT
  - Generalized central path & neighborhood
  - Infeasible iterates
  - Combination of directions as an optimization problem
- Proof of concept: implementable ideas

Thanks!

Fernando R. Villas-Boas

fernandovillasboas@gmail.com