Adaptive Matrix Algebras In Unconstrained Optimization

S. Cipolla, C. Di Fiore

University of Rome "Tor Vergata"
Department of Mathematics

4th International Conference on Matrix Methods in Mathematics and Applications
24th-28th August 2015, Moscow

The Problem

"In 1963 I attended a meeting at Imperial College, London, where most of the participants agreed that the general algorithms of that time for nonlinear optimization calculations were unlikely to be successful if there were more than 10 variables, unless one had an approximation to the solution in the region of convergence of Newton's method. However, because I had studied the report of Davidson that presented the first variable metric algorithm, I already had a computer program that would calculate least values of functions of up to 100 variables using only function values and first derivatives."

M. J. D. Powell

$f : \mathbb{R}^n \to \mathbb{R}$ lower bounded,

find $\boldsymbol{x}_*$ such that

$$f(\boldsymbol{x}_*) = \min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}).$$

Matrix Algebras [1980-2001]

Let U be a unitary matrix, let us define

$$\mathcal{L} = \{Ud(\mathbf{z})U^H \ : \ \mathbf{z} \in \mathbb{C}^n\} = \text{sd}\, U, \qquad d(\mathbf{z}) = diag(z_1, \dots, z_n).$$

Given $A \in M_n(\mathbb{C})$ let us define

- $\mathcal{L}_A = arg\, \min_{X \in \mathcal{L}} ||X - A||_F$, where $||A||_F = \sum_{r,t=1}^n \overline{a}_{rt} a_{rt}$;

Properties $\mathcal{L}_A$

- $\mathcal{L}_A$ well defined because $\mathcal{L}$ is a closed subspace of $\mathbb{C}^{n \times n}$ (Hilbert's Projection Theorem);
- $\mathcal{L}_A = Ud(\mathbf{z}_A)U^H$ where $[\mathbf{z}_A]_i = [U^H A U]_{ii}, \ i = 1, \dots, n$;
- $A \in \mathbb{R}^{n \times n}$, $U \in \mathbb{R}^{n \times n}$ $(U^H = U^T) \ \Rightarrow \mathcal{L}_A \in \mathbb{R}^{n \times n}$;
- A S.P.D (Real Symmetric Positive Definite), $U \in \mathbb{R}^{n \times n}$ $(U^H = U^T) \ \Rightarrow \mathcal{L}_A$ S.P.D;
- $tr\mathcal{L}_A = \sum_i [\mathbf{z}_A]_i = \text{tr}\, A$;
- $det\mathcal{L}_A = \prod_i [\mathbf{z}_A]_i \geq det A$.

$\chi(M)$ number of FLOPS sufficient to perform matrix-vector product Mx, $\mathbf{x} \in \mathbb{C}^n$.

If $L \in \mathcal{L} = sdU$, then $\chi(L) = \chi(U^T) + \chi(U) + n$.

- $\chi(U) = O(n) \Longrightarrow \chi(L) = O(n)$ for all $L \in \mathcal{L}$.

Generalized quasi-Newton: Algorithm Structure

Algorithm 0.1: Generalized Quasi-Newton [2003]

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$;
$\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$;
B_0 S.P.D., $\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;
for $k = 0, 1 \ldots$ **do**
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$;
  $\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$;
  $\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$;
  $B_{k+1} = \Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k)$;
  $\begin{cases} \text{Define } \tilde{B}_{k+1} \ S.P.D \Rightarrow \mathbf{d}_{k+1} = -\tilde{B}_{k+1}^{-1}\mathbf{g}_{k+1} & \mathcal{NS}; \\ \mathbf{d}_{k+1} = -B_{k+1}^{-1}\mathbf{g}_{k+1} \Rightarrow \text{Define } \tilde{B}_{k+1} \ S.P.D & \mathcal{S}; \end{cases}$

- $\Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k) = \tilde{B}_k + \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}\mathbf{y}_k\mathbf{y}_k^T - \frac{1}{\mathbf{s}_k^T \tilde{B}_k \mathbf{s}_k}\tilde{B}_k \mathbf{s}_k \mathbf{s}_k^T \tilde{B}_k$ is the generalized *BFGS-type* updating formula;

Remarks

- $\tilde{B}_k$ is a S.P.D. approx. of B_k;
- if $\tilde{B}_k = B_k$ for all $k = 0, 1, \ldots$ we obtain classical *BFGS* method;
- the $\mathcal{NS}$ algorithm and $\mathcal{S}$ algorithms generate sequences $\{\mathbf{x}_k\}_{k\in\mathbb{N}}, \{\mathbf{g}_k\}_{k\in\mathbb{N}}, \{B_k\}_{k\in\mathbb{N}}$ COMPLETELY DIFFERENT!

Generalized quasi-Newton : Properties

Algorithm 0.2: Generalized q-N

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$;
$\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$;
B_0 S.P.D, $\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;

for $k = 0, 1 \ldots$ **do**

- $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$;
- $\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$;
- $\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$;
- $B_{k+1} = \Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k)$;
- $$\mathbf{d}_{k+1} = \begin{cases} -\tilde{B}_{k+1}^{-1}\mathbf{g}_{k+1} & \mathcal{NS}; \\ -B_{k+1}^{-1}\mathbf{g}_{k+1} & \mathcal{S}; \end{cases}$$

Properties

- $\Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k)\mathbf{s}_k = \mathbf{y}_k$
  $B_{k+1}\mathbf{s}_k = \mathbf{y}_k \rightarrow \mathcal{S}$ecant Algorithm;
  $\tilde{B}_{k+1}\mathbf{s}_k \neq \mathbf{y}_k \rightarrow \mathcal{N}$on $\mathcal{S}$ecant Algorithm;
- $\mathbf{g}_k^T \mathbf{d}_k < 0$ and λ_k such that $(0 < c_1 < c_2 < 1)$:
  $f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + c_1 \lambda_k \mathbf{g}_k^T \mathbf{d}_k$
  $\nabla f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) \geq c_2 \mathbf{g}_k^T \mathbf{d}_k$
  $\Downarrow$
  $\mathbf{s}_k^T \mathbf{y}_k > 0$ and $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$.
- $\mathbf{s}_k^T \mathbf{y}_k > 0$ and $\tilde{B}_k$ S.P.D. $\Rightarrow \Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k)$ S.P.D;

$$\begin{cases} \tilde{B}_{k+1} \text{ S.P.D. } \Rightarrow \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} < 0 \ (\mathcal{NS}); \\ B_{k+1} = \Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k) \text{ S.P.D. } \Rightarrow \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} < 0 \ (\mathcal{S}). \end{cases}$$

Algorithm 0.3: Generalized q-N

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$;
$\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$;
B_0 S.P.D, $\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;

```
for k = 0, 1 . . . do
    x_{k+1} = x_k + λ_k d_k ;
    s_k = x_{k+1} − x_k ;
    y_k = g_{k+1} − g_k ;
    B_{k+1} = Φ(B̃_k, s_k, y_k);
    d_{k+1} = { −B̃_{k+1}^{-1} g_{k+1}   NS;
              { −B_{k+1}^{-1} g_{k+1}    S;
```

$$\mathbf{d}_{k+1} = \begin{cases} -\tilde{B}_{k+1}^{-1}\mathbf{g}_{k+1} & \mathcal{NS}; \\ -B_{k+1}^{-1}\mathbf{g}_{k+1} & \mathcal{S}; \end{cases}$$

- $B_{k+1}^{-1} = \Psi(\tilde{B}_k^{-1}, \mathbf{s}_k, \mathbf{y}_k) = (I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k})^T \tilde{B}_k^{-1} (I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}$.

Complexity

- if $\tilde{B}_k = B_k$ for all $k = 0, 1 \ldots$ we obtain *BFGS* and its complexity is :
  $O(n^2)$ FLOPS per step;
  $O(n^2)$ memory allocations;
- if $\tilde{B}_k \neq B_k$ algorithm's complexity is :
  ■ Time Complexity per Step :
  - number of FLOPS sufficient to calculate $\tilde{B}_k^{-1}$ where $\tilde{B}_k$ is an approximation of B_k;
  - number of FLOPS sufficient to multiply the matrix $\tilde{B}_k^{-1}$ by a vector;
  - $O(n)$ more FLOPS ;

  ■ Space Complexity :
  - number of memory allocation sufficient to store $\tilde{B}_k^{-1}$;
  - $O(n)$ more memory allocation.

Global convergence of generalized $\mathcal{QNNS}$

Algorithm 0.4: $\mathcal{QNNS}$

Data: $\mathbf{x}_0 \in \mathbb{R}^n$;
$\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$, B_0 S.P.D,;
$\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;
for $k = 0, 1 \ldots$ **do**
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$;
  $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$;
  $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$;
  $B_{k+1} = \Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k)$;
  $\mathbf{d}_{k+1} = -\tilde{B}_{k+1}^{-1} \mathbf{g}_{k+1}$;

$\mathcal{N}$on $\mathcal{S}$ecant Global Convergence [2003]

If $\tilde{B}_k$ is such that

$$\begin{cases} trB_k \geq tr\tilde{B}_k \\ detB_k \leq det\tilde{B}_k \end{cases} \tag{1}$$

and there exists $M > 0$ such that

$$\frac{||\mathbf{y}_k||^2}{\mathbf{y}_k^T \mathbf{s}_k} \leq M, \tag{2}$$

then

$$\liminf ||\mathbf{g}_k|| = 0.$$

NOTE 1: (1) is verified if $\tilde{B}_k = \mathcal{L}_{B_k}$ for some $\mathcal{L} =$ sd U.
NOTE 2: (2) is verified if f is convex.

Global convergence of $\mathcal{L}^{(k)}\mathcal{QNNS}$ ("pure projections")

Algorithm 0.5: $\mathcal{L}^{(k)}\mathcal{QNNS}$

Data: $\mathbf{x}_0 \in \mathbb{R}^n$;
$\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$, $\mathcal{L}^{(0)}$;
B_0 S.D.P, $\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;

for $k = 0, 1 \dots$ **do**
- $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$;
- $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$;
- $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$;
- $B_{k+1} = \Phi(\mathcal{L}^{(k)}_{B_k}, \mathbf{s}_k, \mathbf{y}_k)$;
- Define $\mathcal{L}^{(k+1)}$;
- $\mathbf{d}_{k+1} = -[\mathcal{L}^{(k+1)}_{B_{k+1}}]^{-1}\mathbf{g}_{k+1}$;

$$\mathcal{L}^{(k+1)} = \{U_{k+1} d(\mathbf{z}) U_{k+1}^H \;:\; \mathbf{z} \in \mathbb{C}^n\}$$

$$\Downarrow$$

$$\mathrm{tr}\, B_{k+1} = \mathrm{tr}\, \mathcal{L}^{(k+1)}_{B_{k+1}}$$

$$\det B_{k+1} \leq \det \mathcal{L}^{(k+1)}_{B_{k+1}}$$

Remark

The choice $\mathcal{L}^{(k)} \equiv \mathcal{L}$ *for* $k = 0, 1, \dots$ *is allowed!* ($\mathcal{LQNNS}$)

$\mathcal{L}^{(k)}\mathcal{QNNS}$ and $\mathcal{LQNNS}$ are CONVERGENT but NOT EFFICIENT [2003,2015]

Global Convergence Generalized $\mathcal{QNS}$

Algorithm 0.6: $\mathcal{QNS}$

Data: $\mathbf{x}_0 \in \mathbb{R}^n$;
$\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$, B_0 S.D.P,;
$\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;
for $k = 0, 1 \dots$ **do**
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$;
  $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$;
  $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$;
  $B_{k+1} = \Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k)$;
  $\mathbf{d}_{k+1} = -B_{k+1}^{-1} \mathbf{g}_{k+1}$;

Global Convergence $\mathcal{S}$ecant [2015]

If $\tilde{B}_k$ is such that

$$trB_k \geq tr\tilde{B}_k, \qquad detB_k \leq det\tilde{B}_k$$

$$\frac{||B_k \mathbf{s}_k||^2}{(\mathbf{s}_k^T B_k \mathbf{s}_k)^2} \leq \frac{||\tilde{B}_k \mathbf{s}_k||^2}{(\mathbf{s}_k^T \tilde{B}_k \mathbf{s}_k)^2} \qquad (*)$$

and there exists $M > 0$ such that $\frac{||\mathbf{y}_k||^2}{\mathbf{y}_k^T \mathbf{s}_k} \leq M$, then

$$\liminf ||\mathbf{g}_k|| = 0.$$

$$B_k \mathbf{s}_k = \sigma \tilde{B}_k \mathbf{s}_k \quad \Longleftrightarrow \quad -\tilde{B}_k^{-1} \mathbf{g}_k = \sigma \mathbf{d}_k \quad \Longrightarrow \quad (*)$$

$$\sigma = ?$$

One step analysis of Generalized QN$\mathcal{S}$ self-correction properties [1987-2015]

$$\overline{m}\|\mathbf{z}\|^2 \leq \mathbf{z}^T G(\mathbf{x})\mathbf{z} \leq \overline{M}\|\mathbf{z}\|^2 \qquad \forall\ \mathbf{x} \in \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$$

$$\Downarrow$$

$$\overline{G}\mathbf{s}_k = \mathbf{y}_k, \text{ where } \overline{G} = \int_0^1 G(\mathbf{x}_k + \tau \mathbf{s}_k)d\tau$$

$B_{k+1} = \Phi(B_k, \mathbf{s}_k, \mathbf{y}_k)$

$$\text{tr}\, B_{k+1} = \text{tr}\, B_k - \frac{\|B_k\mathbf{s}_k\|^2}{\mathbf{s}_k^T B_k \mathbf{s}_k} + \frac{\|\mathbf{y}_k\|^2}{\mathbf{y}_k^T \mathbf{s}_k}$$

$$\det(B_{k+1}) = \det(B_k)\frac{\mathbf{s}_k^T(\overline{G}\mathbf{s}_k)}{\mathbf{s}_k^T B_k \mathbf{s}_k}$$

$B_{k+1} = \Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k)$

$$\text{tr}\, B_{k+1} = \text{tr}\, \tilde{B}_k - \frac{\|\tilde{B}_k\mathbf{s}_k\|^2}{\mathbf{s}_k^T \tilde{B}_k \mathbf{s}_k} + \frac{\|\mathbf{y}_k\|^2}{\mathbf{y}_k^T \mathbf{s}_k}$$

$$\det(B_{k+1}) = \det(\tilde{B}_k)\frac{\mathbf{s}_k^T(\overline{G}\mathbf{s}_k)}{\mathbf{s}_k^T \tilde{B}_k \mathbf{s}_k}$$

$$\Downarrow$$

If $B_k\mathbf{s}_k = \sigma\tilde{B}_k\mathbf{s}_k,\ \text{tr}\,\tilde{B}_k = \text{tr}\, B_k,\ \det\tilde{B}_k \geq \det B_k$

$$\text{tr}\, B_{k+1} = \text{tr}\, B_k - \frac{1}{\sigma}\frac{\|B_k\mathbf{s}_k\|^2}{\mathbf{s}_k^T B_k \mathbf{s}_k} + \frac{\|\mathbf{y}_k\|^2}{\mathbf{y}_k^T \mathbf{s}_k}, \qquad \det(B_{k+1}) = \sigma \det(\tilde{B}_k)\frac{\mathbf{s}_k^T \overline{G}\mathbf{s}_k}{\mathbf{s}_k^T B_k \mathbf{s}_k}$$

$\sigma = 1 \Rightarrow$ self-correction properties analogous to BFGS!

Global convergence $\mathcal{L}^{(k)}QN\mathcal{S}$, $\tilde{B}_k =$"pure" projection

At each step impose $\dfrac{||B_k \mathbf{s}_k||^2}{(\mathbf{s}_k^T B_k \mathbf{s}_k)^2} = \dfrac{||\tilde{B}_k \mathbf{s}_k||^2}{(\mathbf{s}_k^T \tilde{B}_k \mathbf{s}_k)^2}$

Algorithm 0.7: $\mathcal{L}^{(k)}QN\mathcal{S}$

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$;
$\boldsymbol{g}_0 = \nabla f(\boldsymbol{x}_0)$, $\mathcal{L}^{(0)} =$ sd U_0;
B_0 S.P.D, $\mathbf{d}_0 \in \mathbb{R}^n$, $\mathbf{d}_0^T \mathbf{g}_0 < 0$;
for $k = 0, 1 \ldots$ **do**
  $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$;
  $\boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$;
  $\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$;
  $B_{k+1} = \Phi(\mathcal{L}^{(k)}_{B_k}, \boldsymbol{s}_k, \boldsymbol{y}_k)$;
  $\mathbf{d}_{k+1} = -B_{k+1}^{-1} \mathbf{g}_{k+1}$;
  Choose $\mathcal{L}^{(k+1)}$;

Choose $\mathcal{L}^{(k+1)} \rightarrow$ Totally Non Linear Problem

To guarantee the convergence:
Find U_{k+1} such that

$$B_{k+1}\mathbf{s}_{k+1} = \sigma \mathcal{L}^{(k+1)}_{B_{k+1}} \mathbf{s}_{k+1},$$

where

$$\mathcal{L}^{(k+1)}_{B_{k+1}} = U_{k+1} d(\mathbf{z}_{k+1}) U_{k+1}^T$$

and

$$[\mathbf{z}_{k+1}]_i = [U_{k+1}^T B_{k+1} U_{k+1}]_{ii} > 0, \ \ i = 1, \ldots, n \, .$$

Global convergence $\mathcal{L}^{(k)}QNS$, $\tilde{B}_k =$"hybrid" projection

The matrix $\tilde{B}_{k+1}$ approximation of $B_{k+1} = \Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k)$ must be S.P.D. and in $\mathcal{L}^{(k+1)}$, i.e must have the following **structure** :

$$\tilde{B}_{k+1} = U_{k+1} d(\mathbf{z}_{k+1}) U_{k+1}^T, \quad U_{k+1} \text{ unitary, } \mathbf{z}_{k+1} > 0, \quad \chi(U_{k+1}) << n^2.$$

Which kind of structure should have $\mathcal{L}^{(k+1)} =$ sd U_{k+1} and which kind of spectrum $\mathbf{z}_{k+1}$ should have $\tilde{B}_{k+1}$ in order to guarantee convergence?

Algorithm 0.8: Hybrid $\mathcal{L}^{(k)}QNS$

```
Data: ...
for k = 0, 1 ... do
    x_{k+1} = x_k + λ_k d_k ;
    s_k = x_{k+1} − x_k;
    y_k = g_{k+1} − g_k;
    B_{k+1} = Φ(B̃_k, s_k, y_k);
    d_{k+1} = −B_{k+1}^{-1} g_{k+1};
    Define z_{k+1} > 0;
    Choose L^{(k+1)} (i.e choose U_{k+1});
    Define B̃_{k+1} = U_{k+1} d(z_{k+1}) U_{k+1}^T;
```

Partially Non Linear Problem

To guarantee the convergence *it is sufficient*

Given $\mathbf{z} \in \mathbb{R}^n$, $\mathbf{z} = \mathbf{z}_{k+1} > 0$, such that

$\det B_{k+1} \leq \prod z_i$, $\operatorname{tr} B_{k+1} \geq \sum z_i$

Find U_{k+1} unitary such that

$B_{k+1}\mathbf{s}_{k+1} = \sigma U_{k+1} d(\mathbf{z}) U_{k+1}^T \mathbf{s}_{k+1}$

EXAMPLE: Try for

$[\mathbf{z}_{k+1}]_i = [U_k^T B_{k+1} U_k]_{ii}$, i.e. $\mathbf{z}_{k+1} = \lambda(\mathcal{L}^{(k)}_{B_{k+1}})$.

Existence of the solution for PNLP (using σ as a parameter) [2015]

Given $\mathbf{z} > 0$, exists U_{k+1} unitary and $\sigma_{k+1} > 0$ such that

$$B_{k+1}\mathbf{s}_{k+1} = \sigma_{k+1} U_{k+1} d(\mathbf{z}) U_{k+1}^T \mathbf{s}_{k+1}$$

if and only if the following Kantorovich condition holds

$$\frac{4 z_m z_M}{(z_m + z_M)^2} \leq \frac{(\mathbf{s}_{k+1}^T(-\mathbf{g}_{k+1}))^2}{||\mathbf{s}_{k+1}||^2 ||\mathbf{g}_{k+1}||^2}.$$

"$\Leftarrow$": Starting from $\mathbf{z} > 0$ such that hypothesis are fulfilled, we build explicitly

$$U_{k+1} = \mathbf{dHc}(\mathbf{z}) = H(\mathbf{w}_{k+1}) H(\mathbf{v}_{k+1}),$$

where $H(\mathbf{x})$ is the Householder matrix $I - \frac{2}{||\mathbf{x}||^2}\mathbf{x}\mathbf{x}^T$.
Let us denote the corresponding algebra

$$\mathcal{L}^{(k+1)} = \text{sd } U_{k+1} =: [2Ho]^{(k+1)}$$

Observe that $\chi(H(\mathbf{x})) = O(n)$ for all $\mathbf{x} \in \mathbb{R}^n \implies \chi(L) = O(n)$ for all $L \in [2Ho]^{(k+1)}$

Existence of the solution for PNLP with $\sigma = 1$

Given $\mathbf{z} > 0$ such that

$$z_m < \frac{\mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1}}{\|\mathbf{s}_{k+1}\|^2} < z_M,$$

$$\|B_{k+1}\mathbf{s}_{k+1}\|^2 - (z_m + z_M)\mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1} + z_M z_m \|\mathbf{s}_{k+1}\|^2 \leq 0$$

and exists $\bar{j} \in \{1, \ldots, n\} \backslash \{m, M\}$ such that

$$z_{\bar{j}} \in [\frac{\mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1} z_M - \|B_{k+1}\mathbf{s}_{k+1}\|^2}{z_M\|\mathbf{s}_{k+1}\|^2 - \mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1}}, \frac{\mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1} z_m - \|B_{k+1}\mathbf{s}_{k+1}\|^2}{z_m\|\mathbf{s}\|^2 - \mathbf{s}_{k+1}^T B_{k+1}\mathbf{s}_{k+1}}] =: [\tilde{\theta}_{\mathbf{s}}, \tilde{\beta}_{\mathbf{s}}],$$

(we will write $\mathcal{P}(z_m, z_M) = True$) then there exists a unitary U_{k+1} such that

$$B_{k+1}\mathbf{s}_{k+1} = U_{k+1} d(\mathbf{z}) U_{k+1}^T \mathbf{s}_{k+1}.$$

"$\Leftarrow$": Starting from $\mathbf{z} > 0$ such that $\mathcal{P}(z_m, z_M) = T$, we build explicitly

$$U_{k+1} = \mathbf{dHc}(\mathbf{z}) = H(\mathbf{w}_{k+1})H(\mathbf{v}_{k+1}),$$

where $H(\mathbf{x})$ is the Householder matrix $I - \frac{2}{||\mathbf{x}||^2}\mathbf{x}\mathbf{x}^T$.
In this case let us denote the corresponding algebra

$$\mathcal{L}^{(k+1)} = \text{sd } U_{k+1} =: [2Ho]^{(k+1)}$$

Observe that $\chi(H(\mathbf{x})) = O(n)$ for all $\mathbf{x} \in \mathbb{R}^n \implies \chi(L) = O(n)$ for all $L \in [2Ho]^{(k+1)}$

Algorithm 0.9: Hybrid $\mathcal{L}^{(k)}QN\mathcal{S}$

Data: ...

1 **for** $k = 0, 1 \dots$ **do**

2 $\quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k$;

3 $\quad \boldsymbol{s}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$;

4 $\quad \boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$;

5 $\quad B_{k+1} = \Phi(\tilde{B}_k, \boldsymbol{s}_k, \boldsymbol{y}_k)$;

6 $\quad \boldsymbol{d}_{k+1} = -B_{k+1}^{-1} \boldsymbol{g}_{k+1}$;

7 $\quad$ *Consider* $[2Ho]^{(k)}_{B_{k+1}}$;

8 $\quad$ *Compute* $\mathbf{z}_{k+1} = \lambda([2Ho]^{(k)}_{B_{k+1}})$;

9 $\quad$ **if** $\mathcal{P}([\mathbf{z}_{k+1}]_m, [\mathbf{z}_{k+1}]_M) = T$ **then**

10 $\quad\quad U_{k+1} = \mathbf{dHc}(\mathbf{z}_{k+1})$;

11 $\quad\quad \tilde{B}_{k+1} = U_{k+1} d(\mathbf{z}_{k+1}) U_{k+1}^T$;

12 $\quad\quad [2Ho]^{(k+1)} = \text{sd } U_{k+1}$;

13 $\quad$ **else**

14 $\quad\quad \bar{\mathbf{z}}_{k+1} = \mathbf{SC}(\mathbf{z}_{k+1})$;

15 $\quad\quad \mathbf{z}_{k+1} := \bar{\mathbf{z}}_{k+1}$;

16 $\quad\quad U_{k+1} = \mathbf{dHc}(\mathbf{z}_{k+1})$;

17 $\quad\quad \tilde{B}_{k+1} = U_{k+1} d(\mathbf{z}_{k+1}) U_{k+1}^T$;

18 $\quad\quad [2Ho]^{(k+1)} = \text{sd } U_{k+1}$;

Complexity

Set $[2Ho]^{(k)} = \text{sd } U_k$, where $U_k = H(\mathbf{w}_k)H(\mathbf{v}_k)$.

■ $O(n)$ memory allocations are sufficient for implementation.

■ Line (6): $O(n)$ FLOPS

- Invert a matrix in $[2Ho]^{(k)}$;
- Multiply a matrix in $[2Ho]^{(k)}$ by a vector;

■ Line (8) : $O(n)$ FLOPS via a simple eigenvalue updating formula

$$\mathbf{z}_k \to \mathbf{z}_{k+1} = \lambda([2Ho]^{(k)}_{B_{k+1}}).$$

■ Line (10) o (15): $O(n)$ FLOPS for the construction of $U = dHc(z)$.

■ If $\mathcal{P}(z_m, z_M) = False$? Which is the computational cost of **SC**?

If $\mathcal{P}([\mathbf{z}_{k+1}]_m, [\mathbf{z}_{k+1}]_M) = \mathit{False}$? Spectral correction: the strategy

Given $z_i = (U_k^H B_{k+1} U_k)_{ii} > 0$ such that

$$\mathcal{P}(z_m, z_M) = F,$$

we need to produce a correction $\bar{\mathbf{z}}$ of $\mathbf{z}$

$$\bar{\mathbf{z}} := \mathbf{SC}(\mathbf{z}),$$

such that

- $\mathcal{P}(\bar{z}_m, \bar{z}_M) = T$;
- $\sum_{i=1}^n \bar{z}_i \leq \operatorname{tr} B_{k+1}$;
- $\prod_{i=1}^n \bar{z}_i \geq \det B_{k+1}$.

The last two conditions hold any time

$$\tilde{z}_i = (\tilde{V}^H B_{k+1} \tilde{V})_{ii}$$

for some $\tilde{V}$ unitary

Spectral correction: the theoretical framework for $\sigma = 1$

Theorem

Let be B a S.P.D. matrix and $\mathbf{s} \in \mathbb{R}^n$ a given vector. Then:

$$\|B\mathbf{s}\|^2 - (\lambda_m + \lambda_M)\mathbf{s}^T B\mathbf{s} + \lambda_M \lambda_m \|\mathbf{s}\|^2 \leq 0$$

Assumption ($B\mathbf{x}_j = \lambda_j \mathbf{x}_j$ where λ_j are all simple)

$$\frac{\mathbf{s}}{\|\mathbf{s}\|} \neq \mathbf{x}_j \text{ for all } j \in \{1, \ldots, n\}.$$

Theorem

$\forall \ \mathbf{s} \in \mathbb{R}^n$

$$\frac{\mathbf{s}^T B\mathbf{s}}{\|\mathbf{s}\|^2} \in [\theta_{\mathbf{s}}, \beta_{\mathbf{s}}] := [\frac{\mathbf{s}^T B\mathbf{s}\lambda_M - \|B\mathbf{s}\|^2}{\lambda_M \|\mathbf{s}\|^2 - \mathbf{s}^T B\mathbf{s}}, \frac{\mathbf{s}^T B\mathbf{s}\lambda_m - \|B\mathbf{s}\|^2}{\lambda_m \|\mathbf{s}\|^2 - \mathbf{s}^T B\mathbf{s}}],$$

$$\theta_{\mathbf{s}} \leq \beta_{\mathbf{s}}.$$

Theorem

$$\beta_{\mathbf{s}} \geq \lambda_{r(m)} \ \textit{and} \ \theta_{\mathbf{s}} \leq \lambda_{l(M)}.$$

Theorem

Let us suppose to have the following four eigenpairs of B_{k+1}:

$$(\lambda_m, \mathbf{x}_m), \;\; (\lambda_{r(m)}, \mathbf{x}_{r(m)}), \;\; (\lambda_{l(M)}, \mathbf{x}_{l(M)}) \quad \text{and } (\lambda_M, \mathbf{x}_M).$$

Then there exists a unitary $\overline{V}$ such that defining

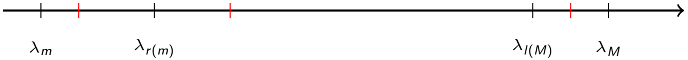
$$\overline{z}_i = (\overline{V}^H B_{k+1} \overline{V})_{ii} \quad \text{for} \quad i = 1, \dots, n,$$

it holds that

$$\mathcal{P}(\overline{z}_m, \overline{z}_M) = T.$$

- $\frac{\mathbf{s}_{k+1}^T B_{k+1} \mathbf{s}_{k+1}}{\|\mathbf{s}_{k+1}\|^2} \in (\lambda_m, \lambda_{r(m)}) \quad \Longrightarrow \quad \overline{V}\mathbf{e}_m = \mathbf{x}_m, \;\; \overline{V}\mathbf{e}_{r(m)} = \mathbf{x}_{r(m)}, \;\; \overline{V}\mathbf{e}_M = \mathbf{x}_M$
- $\frac{\mathbf{s}_{k+1}^T B_{k+1} \mathbf{s}_{k+1}}{\|\mathbf{s}_{k+1}\|^2} \in (\lambda_l(M), \lambda_M) \quad \Longrightarrow \quad \overline{V}\mathbf{e}_m = \mathbf{x}_m, \;\; \overline{V}\mathbf{e}_{l(M)} = \mathbf{x}_{l(M)}, \;\; \overline{V}\mathbf{e}_M = \mathbf{x}_M$
- $\frac{\mathbf{s}_{k+1}^T B_{k+1} \mathbf{s}_{k+1}}{\|\mathbf{s}_{k+1}\|^2} \in (\lambda_r(m), \lambda_{l(M)}) \quad \Longrightarrow \quad \overline{V}\mathbf{e}_m = \mathbf{x}_m, \;\; \overline{V}\mathbf{e}_l = \mathbf{x}_{\mathbf{s}_{k+1}} \;\; (l \neq m, M) \;, \;\; \overline{V}\mathbf{e}_M = \mathbf{x}_M$

where $\mathbf{x}_{\mathbf{s}_{k+1}}$ is such that $\|\mathbf{x}_{\mathbf{s}_{k+1}}\| = 1$, $\mathbf{x}_{\mathbf{s}_{k+1}}^T B_{k+1} \mathbf{x}_{\mathbf{s}_{k+1}} = \frac{\mathbf{s}_{k+1}^T B_{k+1} \mathbf{s}_{k+1}}{\|\mathbf{s}_{k+1}\|^2}$ and $\mathbf{x}_{\mathbf{s}_{k+1}} \in < \mathbf{x}_{r(m)}, \mathbf{x}_{l(M)} >$.



Algorithm 0.10: Coupled Subspace Iteration

Data: $\mathbf{z}_0 := \mathbf{z}, \quad \theta_{\mathbf{s},0} = \frac{\mathbf{s}^T B\mathbf{s} z_M - \|B\mathbf{s}\|^2}{z_M \|\mathbf{s}\|^2 - \mathbf{s}^T B\mathbf{s}}, \quad \beta_{\mathbf{s},0} = \frac{\mathbf{s}^T B\mathbf{s} z_m - \|B\mathbf{s}\|^2}{z_m \|\mathbf{s}\|^2 - \mathbf{s}^T B\mathbf{s}}$

$S_B^{(0)} = [U\mathbf{e}_M, U\mathbf{e}_t], \quad S_{B-1}^{(0)} = [U\mathbf{e}_b, U\mathbf{e}_m]$ where $t, b \neq M, m$

$z_M^0 = z_M, \quad z_t^0 = z_t \quad z_b^0 = z_b, \quad z_m^0 = z_m$

```
i=0;
while STOP CONDITION do
    Z_B^(i+1) = B S_B^(i) ;
    Z_B^(i+1) =: Q_B^(i+1) R_B^(i+1) (QR decomposition);
    Find P_B^(i+1) such that
    [Q_B^(i+1) P_B^(i+1)]^H B Q_B^(i+1) P_B^(i+1) = diag(z_M^(i+1), z_t^(i+1));
    S_B^(i+1) := Q_B^(i+1) P_B^(i+1);
    Z_{B-1}^(i+1) = B^{-1} S_{B-1}^(i) ;
    Z_{B-1}^(i+1) =: Q_{B-1}^(i+1) R_{B-1}^(i+1) (QR decomposition);
    Find P_{B-1}^(i+1) such that
    [Q_{B-1}^(i+1) P_{B-1}^(i+1)]^H B^{-1} Q_{B-1}^(i+1) P_{B-1}^(i+1) = diag(z_b^(i+1), z_m^(i+1));
    S_{B-1}^(i+1) := Q_{B-1}^(i+1) P_{B-1}^(i+1);
    θ_{s,i+1} = (s^T B s z_M^(i+1) − ||Bs||^2) / (z_M^(i+1) ||s||^2 − s^T B s);
    β_{s,i+1} = (s^T B s z_m^(i+1) − ||Bs||^2) / (z_m^(i+1) ||s||^2 − s^T B s);
    i:=i+1 ;
```

$z_i = (U_k^H B_{k+1} U_k)_{ii} > 0$ such that

$$\mathcal{P}(z_m, z_M) = F,$$

Lemma

Algorithm 0.10 produces the mutually orthogonal sequences

$$\{\mathbf{v}_M^i\}_i, \quad \{\mathbf{v}_m^i\}_i, \quad \{\mathbf{v}_t^i\}_i, \quad \{\mathbf{v}_b^i\}_i$$

(columns of the matrices $S_B^{(i)}$, $S_{B-1}^{(i)}$*)*
and the sequences

$$\{z_M^i\}_i, \quad \{z_m^i\}_i, \quad \{z_t^i\}_i \text{ and } \quad \{z_b^i\}_i,$$

such that

$$\lim_{i\to\infty} (\mathbf{v}_M^i, z_M^i) = (\lambda_M, \mathbf{x}_M),$$

$$\lim_{i\to\infty} (\mathbf{v}_m^i, z_m^i) = (1/\lambda_m, \mathbf{x}_m),$$

$$\lim_{i\to\infty} (\mathbf{v}_t^i, z_t^i) = (\lambda_{l(M)}, \mathbf{x}_{l(M)}),$$

$$\lim_{i\to\infty} (\mathbf{v}_b^i, z_b^i) = (1/\lambda_{r(m)}, \mathbf{x}_{r(m)}).$$

[1] C.Di Fiore, P.Zellini, *Matrix algebras in optimal preconditioning*, Linear Algebra and its Applications, 335, 1-54 (2001).

[2] C.Di Fiore, S.Fanelli, F.Lepore, P.Zellini, *Matrix algebras in Quasi-Newton methods for unconstrained minimization*, Numerische Mathematik, 94, 479-500 (2003).

[3] A. Bortoletti, C. Di Fiore, S. Fanelli, P. Zellini, *A new class of quasi-Newtonian methods for optimal learning in MLP-networks*, IEEE Transactions on Neural Networks, 14, 263-273 (2003).

[4] S. Cipolla, C. Di Fiore, F. Tudisco, P. Zellini, *Adaptive Matrix Algebras in unconstrained minimization*, Linear Algebra and its Applications, 471, 544-568, (2015).

[5] R. H. Byrd, J. Nocedal, Y. Yuan, *Convergence of a Class of Quasi-Newton Methods on Convex Problems*, SIAM Journal of Numerical Analysis, 24-5, 1171-1190 (1987).