Sparse optimization techniques for solving multilinear least-squares problems with application to design of filter networks

Spartak Zikrin

Division of Optimization
Department of Mathematics
Linköping University

Joint work with:
Oleg Burdakov, Hans Knutsson, Mats Andersson

Conference on Inverse Problems and Applications, Linköping, Sweden
April 4, 2013

Problem formulation

Linear least-squares problem:

$$\min_{x} \|b - Ax\|^2$$

Multilinear least-squares problem (MLLS):

$$\min_{x_1,\dots,x_L} \|b - (A_1x_1) \circ (A_2x_2) \circ \cdots \circ (A_Lx_L)\|^2$$

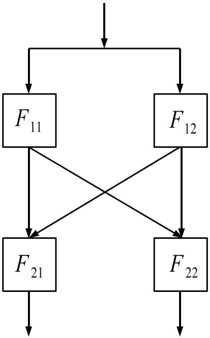
where $u \circ v$ denotes the component-wise product of vectors u and v

Problem origin

- Design of filter networks
  - Medical Image Analysis
- Factor Analysis
- Psychometrics
- Chemometrics

Filter networks

Signal processing time vs fitness to the ideal filter

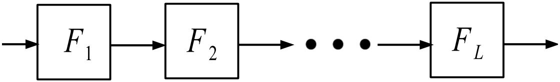


Design of filter networks:

1. network structure
2. sparsity pattern
3. sub-filters coefficients

Sequential filtering

Sequence of L subfilters



Filter design problem with chosen sparsity pattern:

$$\min_{x_1,\dots,x_L} \|b - (A_1x_1) \circ (A_2x_2) \circ \cdots \circ (A_Lx_L)\|^2,$$

where

- $b \in R^m$ is the ideal frequency response.
- Characteristics for subfilter i: A_i, x_i.
- Actual frequency response of the subfilter sequence: $(A_1x_1) \circ (A_2x_2) \circ \cdots \circ (A_Lx_L)$.

Features of MLLS problems

- MLLS is a non-convex large-scale optimization problem.
(In design of filters of dimensionality d, the most typical values are $m \in [10^d, 40^d]$, $L \in [2, 10]$ and $n = n_1 + \cdots + n_L \in [5d, 60d]$.)
- MLLS has a large number of local minima.
- Each local minimizer is singular and non-isolated.
(If $x_1, \ldots, x_L$ is a local minimizer, then $t_1 x_1, \ldots, t_L x_L$ is also a local minimizer for any combination of scalars $t_1, \ldots, t_L$ such that $t_1 \cdot \ldots \cdot t_L = 1$.)
- The Jacobian matrix of $(A_1 x_1) \circ (A_2 x_2) \circ \cdots \circ (A_L x_L)$ is rank deficient (the rank deficiency $\geq L - 1$).

Generation of starting point

Reformulation of MLLS:

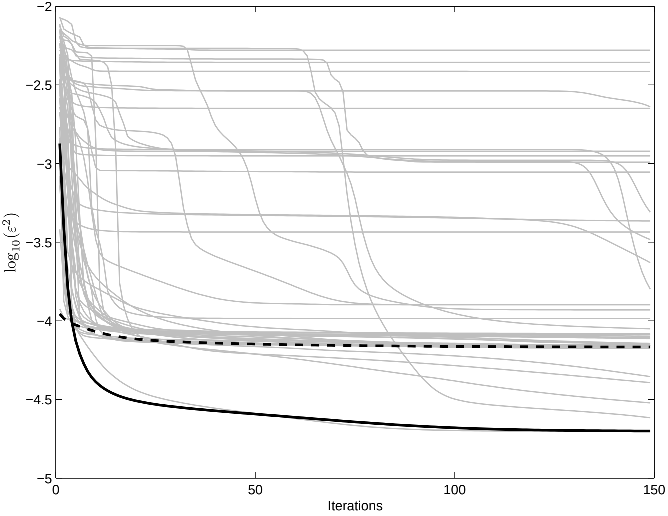
$$\min_{\substack{x_1,\ldots,x_L \\ b_1,\ldots,b_L}} \|b - b_1 \circ \cdots \circ b_L\|^2 \quad \text{s.t.} \quad A_i x_i = b_i, \qquad i = 1,\ldots,L$$

$$\Downarrow$$

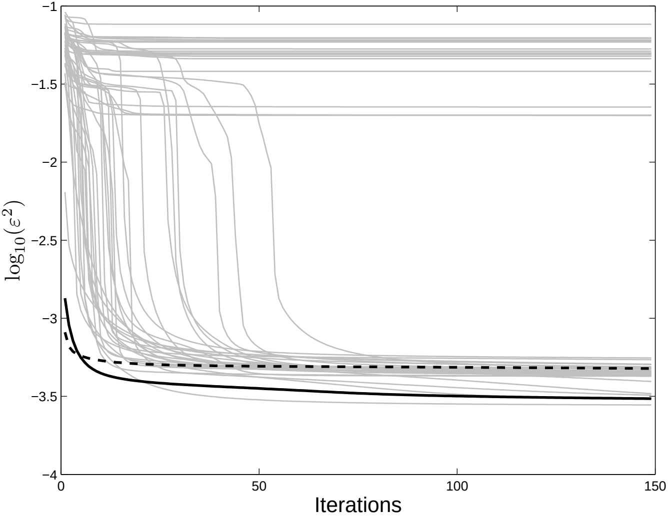
$$\min_{\substack{x_1,\ldots,x_L \\ b_1,\cdots,b_L}} \sum_1^L \|b_i - A_i x_i\|^2 \quad \text{s.t.} \quad b_1 \circ \ldots \circ b_L = b$$

- Conceptually close to MLLS
- Convex objective function
- Efficient global search strategy with easy-to-solve subproblems (half CPU time of one ALS run)
- Use the solution as a starting point for one run of ALS.

Band-pass filter of order 2



High-pass filter of order 2



Subfilter sparsity optimization

Cardinality-constrained MLLS:

$$\min_{\|\bar{x}\|_0 \le n} \|b - (\bar{A}_1\bar{x}_1) \circ (\bar{A}_2\bar{x}_2) \circ \cdots \circ (\bar{A}_L\bar{x}_L)\|^2$$

where $\bar{x} = [\bar{x}_1; \ldots; \bar{x}_L] \in R^{\bar{n}_1 + \cdots + \bar{n}_L}$, $\bar{A}_i \in R^{m \times \bar{n}_i}$, $\bar{n}_i \in [3^d, 15^d]$ and $n \in [5d, 60d]$

Subfilter sparsity optimization

Cardinality-constrained MLLS:

$$\min_{\|\bar{x}\|_0 \le n} \|b - (\bar{A}_1\bar{x}_1) \circ (\bar{A}_2\bar{x}_2) \circ \cdots \circ (\bar{A}_L\bar{x}_L)\|^2$$

where $\bar{x} = [\bar{x}_1; \ldots; \bar{x}_L] \in R^{\bar{n}_1 + \cdots + \bar{n}_L}$, $\bar{A}_i \in R^{m \times \bar{n}_i}$, $\bar{n}_i \in [3^d, 15^d]$ and $n \in [5d, 60d]$

Cardinality-constrained linear least-squares problem:

$$\min_{\|\bar{x}_i\|_0 \le n_i} \|b - \bar{D}_i\bar{A}_i\bar{x}_i\|^2,$$

$$\bar{D}_i = diag\left((\bar{A}_1\bar{x}_1) \circ \cdots \circ (\bar{A}_{i-1}\bar{x}_{i-1}) \circ (\bar{A}_{i+1}\bar{x}_{i+1}) \circ \cdots \circ (\bar{A}_L\bar{x}_L)\right).$$

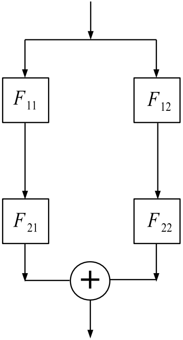
Procedure

```
improvement = true
while improvement do
  for α = 0, 1, ..., α_max do
    for i = 1, ..., L do
      φ_best = φ_c
      x̄_c → D̄_i, n_i
      if n_i − α > 0 then
        GREEDY(b, D̄_i Ā_i, n_i − α) → x̄_r
        ALS(x̄_t) → x̄_r
        for k = 1, ..., i − 1, i + 1, ..., L do
          x̄_r → D̄_k, n_k
          if n_k + α ≤ n̄_k then
            GREEDY(b, D̄_k Ā_k, n_k + α) → x̄_t
            ALS (x̄_t) → x̄_t, φ_t
            if φ_t < φ_best then
              x̄_best = x̄_t
              φ_best = φ_t
              improvement = true
            end if
          end if
        end for
        x̄_c = x̄_best
        φ_c = φ_best
      end if
    end for
  end for
end while
```

Test results

Filter	$\min(\varepsilon_j)^2$	$(\varepsilon_{sp})^2$	ratio	$\Delta\varepsilon^2$	$\min(\varepsilon_j^{new})^2$	time
Low-Pass	3.30e-5	2.80e-5	1.18	15.07 %	7.90e-5	2073.5 s
Band-Pass (BP)	5.77e-6	4.52e-6	1.28	21.65 %	5.43e-6	2293.2 s
High-Pass (HP)	3.02e-4	1.51e-4	1.99	49.85 %	8.13e-4	2289.8 s
BP, 1:st order	5.76e-5	5.76e-5	1	0 %	5.76e-5	749.5 s
HP, 1:st order	1.18e-2	1.18e-2	1	0 %	1.18e-2	747.4 s
BP, 2:nd order	5.93e-6	5.56e-6	1.07	6.16 %	3.13e-6	1522.4 s
BP, 2:nd order (cont.)	3.13e-6	2.70e-6	1.16	13.78 %	4.16e-6	1524.3 s
HP, 2:nd order	2.78e-4	2.53e-4	1.10	8.71 %	3.36e-4	1482.9 s

Design of layered filter network

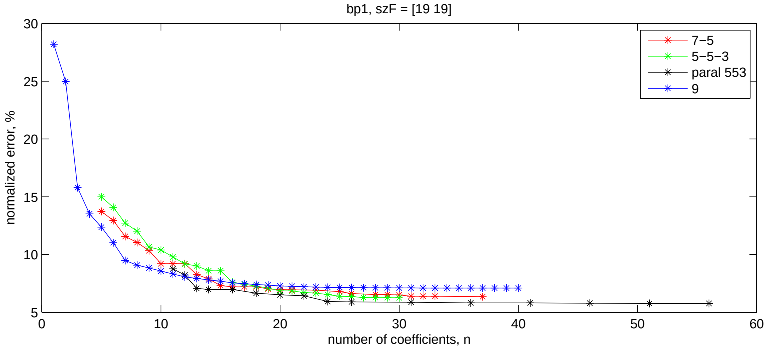


Method	Normalized error
Low-Rank Approximation	11.9%
Expert Filter Design	10.9%
Our Sparse Approximation	4.9%

First layer:

$$\min_{\|\bar{x}_1\|_0 \le \bar{n}_1} \left\| b - \begin{bmatrix} \bar{D}_{21}\bar{A}_{11} & \bar{D}_{22}\bar{A}_{12} \end{bmatrix} \bar{x}_1 \right\|^2, \qquad \bar{x}_1 = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{12} \end{bmatrix}$$

Bi-criteria optimization problem



Conclusions

- faster filter optimization
- faster signal processing time

Future work:

- More efficient solvers for subroutines
- New areas of application

Conclusions

- faster filter optimization
- faster signal processing time

Future work:

- More efficient solvers for subroutines
- New areas of application

Thank you for attention