A Multicore Convex Optimization Algorithm with Applications to Video Restoration

Feriel Abboud[1], *Emilie Chouzenoux*[2,3], *Jean-Christophe Pesquet*[2] *and Hugues Talbot*[2,3]

[1] WITBE, Puteaux, France, [2] CVN, CentraleSupélec, INRIA Galen, Univ. Paris-Saclay, France, [3] LIGM, CNRS, ESIEE, Univ. Paris Est, France

Problem formulation

★ New distributed algorithm for minimizing a sum of non-necessarily differentiable convex functions composed with arbitrary linear operators.

- ✔ **Hypergraph structure** modeling communication with computing units,
- ✔ Primal-dual splitting strategy with **established convergence guarantees**,
- ✔ Use of modern **parallel architecture**,
- ✔ Application to **video restoration** ⇝ Significant speedup.

Minimization problem

Let G a proper lower-semicontinuous convex functions from $\mathbb{R}^N$ to $]-\infty,+\infty]$ and $\widetilde{x}$ a given point of $\mathbb{R}^N$. We aim at computing the proximity operator of G at $\tilde{x}$:

$$\text{Find } \widetilde{x} = \underset{x\in\mathbb{R}^N}{\operatorname{argmin}}\; G(x) + \frac{1}{2}\|x-\widetilde{x}\|^2, \\ = \operatorname{prox}_G(\widetilde{x}).$$

We focus on functions G with the form:

$$(\forall x \in \mathbb{R}^N) \quad G(x) = \textstyle\sum_{j=1}^J g_j(A_j x)$$

- For every $j \in \{1,\dots,J\}$, $g_j : \mathbb{R}^{M_j} \to]-\infty,+\infty]$ is a proper lower-semicontinuous, possibly nonsmooth, convex function,
- A_j is a real-valued matrix of dimension $M_j \times N$,
- $\bigcap_{j=1}^J \operatorname{dom}(g_j \circ A_j)$ is nonempty.

Primal-dual splitting methods:

- Minimization of non smooth convex functions, with established convergence guarantees
- Well suited to functions composed with large-size linear operators (no inversion required)
- Recently combined with block-coordinate approaches, leading to fast convergence speed.

Challenge: Available algorithms rely on a centralized implementation ⇝ Unsuitable for dealing with massive datasets.

Proposed Method

Consensus formulation:

Replace the global variable x by a collection of vectors $(x^j)_{1\leqslant j\leqslant J}$, and ensure the coupling through Λ, a vector subspace of $\mathbb{R}^{NJ}$.

$$\underset{\boldsymbol{x}=(x^j)_{1\leqslant j\leqslant J}\in\Lambda}{\operatorname{argmin}} \quad \textstyle\sum_{j=1}^J g_j(A_j x^j) + \frac{1}{2}\sum_{j=1}^J \omega_j \|x^j - \widetilde{x}\|^2$$

with $(\omega_j)_{1\leqslant j\leqslant J} \in]0,1]^J$ such that $\sum_{j=1}^J \omega_j = 1$.

Hypergraph structure:

- Local nodes $j \in \{1,\dots,J\}$ associated with g_j, processing their own private data.
- Communication only allowed between nodes belonging to the same group $\mathbb{V}_\ell$ with $\ell \in \{1,\dots,L\}$.

⇝ Decomposition of Λ into $(\Lambda_\ell)_{1\leqslant \ell\leqslant L}$ such that:

$$\boldsymbol{x} \in \Lambda \quad \Leftrightarrow \quad (\forall \ell \in \{1,\dots,L\}) \quad (x^j)_{j\in\mathbb{V}_\ell} \in \Lambda_\ell.$$

Distributed implementation:

For every $j \in \{1,\dots,J\}$, assume that:

$$(\forall x^j = ([x^j]_t)_{1\leqslant t\leqslant T} \in \mathbb{R}^N) \quad A_j x^j = \textstyle\sum_{t\in\mathbb{T}_j} \mathcal{A}_{j,t}[x^j]_t$$

with $[x^j]_t$ a block of data of dimension L, T the overall number of blocks (i.e., $N = TL$), and $\mathbb{T}_j \subset \{1,\dots,T\}$ the reduced index subset of the components of vector x^j acting on the operator A_j. Then, the initial problem is equivalent to minimizing:

$$\textstyle\sum_{j=1}^J g_j\Big(\sum_{t\in\mathbb{T}_j} \mathcal{A}_{j,t}[x^j]_t\Big) + \frac{1}{2}\sum_{j=1}^J \sum_{t\in\mathbb{T}_j} \omega_{j,t}\big\|[x^j]_t - [\widetilde{x}]_t\big\|^2$$

Minimization method:

⇝ Dual forward-backward algorithm, accelerated by a variable metric strategy.

At each iteration $n \in \mathbb{N}$, an index $j_n \in \{1,\dots,J+L\}$ is selected:

- If $j_n \leqslant J$, local optimization step on function g_{j_n}.
- If $j_n > J$, synchronization step, averaging operation within the nodes $j \in \mathbb{V}_{j_n - J}$.

⇒ Convergence of the sequence of iterates to the solution.

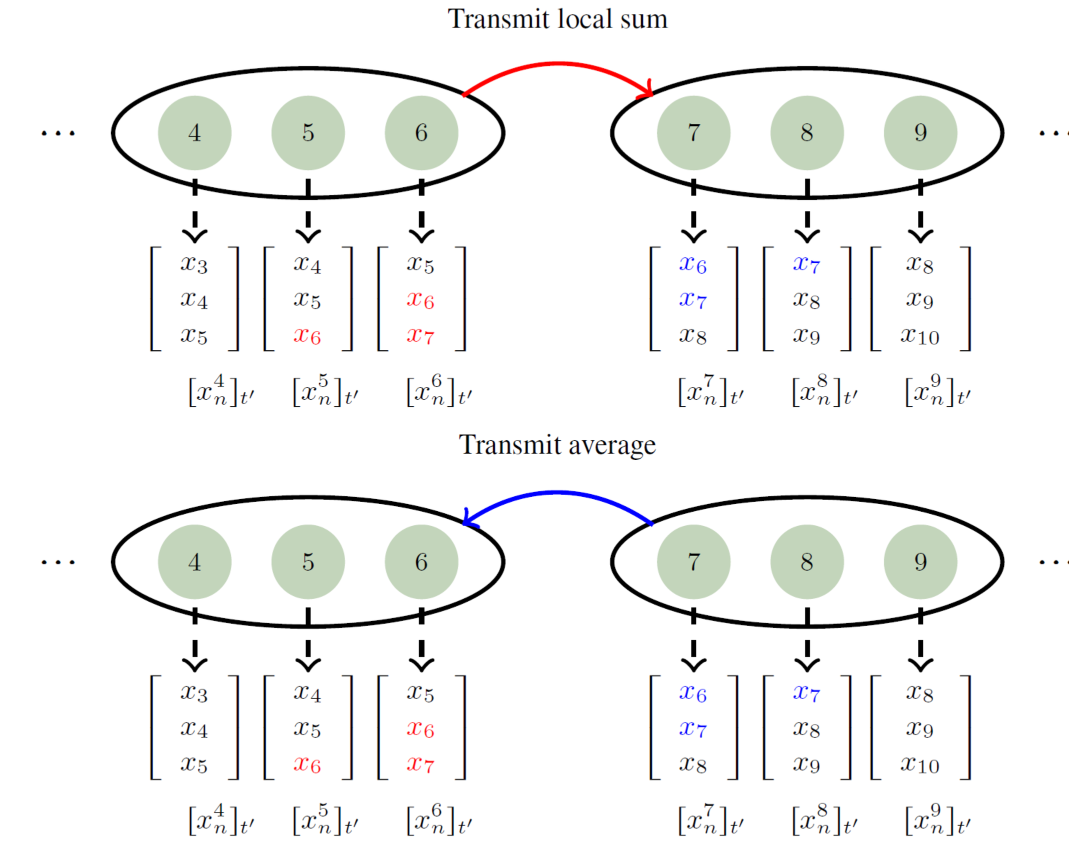
Experimental results

Problem statement:

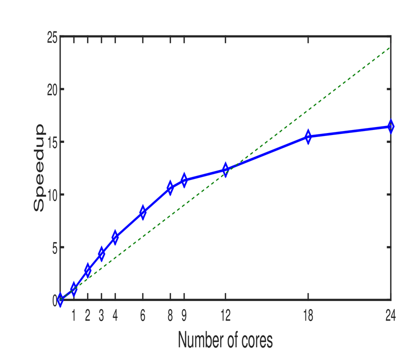
Restoration of an original video sequence with T frames, and L pixels per frame, corrupted by zero-mean white Gaussian noise.

⇝ Minimization of a penalized least squares function, with several regularization terms $(g_t \circ A_t)_{1\leqslant t\leqslant T}$ (TV, range constraint, temporal regularization).

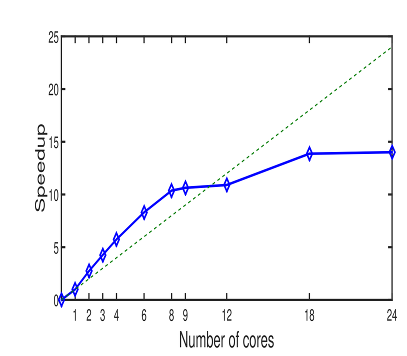
Implementation:



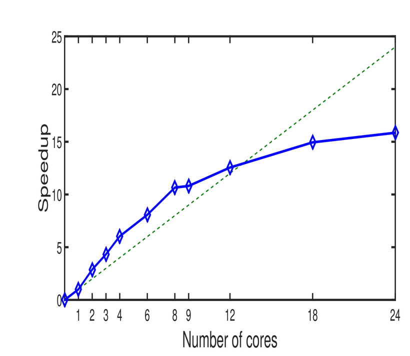
Implemented in the Julia langage with Message Passing Interface
Tests performed on Intel E5 v4 multi-core architecture with $C = 24$ cores.



Foreman
$348 \times 284 \times 72$



Claire
$300 \times 278 \times 32$



Irene
$352 \times 288 \times 72$