Conservative Signal Processing Architectures
For Asynchronous, Distributed Optimization
Part II: Example Systems

Thomas A. Baran and Tarek A. Lahlou
Digital Signal Processing Group
Massachusetts Institute of Technology

Abstract—This paper provides examples of various synchronous and
asynchronous signal processing systems for performing optimization,
utilizing the framework and elements developed in a preceding paper. The
general strategy in that paper was to perform a linear transformation of
stationarity conditions applicable to a class of convex and nonconvex
optimization problems, resulting in algorithms that operate on a linear
superposition of the associated primal and dual decision variables. The
examples in this paper address various specific optimization problems
including the LASSO problem, minimax-optimal filter design, the decen-
tralized training of a support vector machine classifier, and sparse filter
design for acoustic equalization. Where appropriate, multiple algorithms
for solving the same optimization problem are presented, illustrating the
use of the underlying framework in designing a variety of distinct classes
of algorithms. The examples are accompanied by numerical simulation
and a discussion of convergence.

Index Terms—Asynchronous optimization, distributed optimization,
conservation

I. INTRODUCTION

This paper presents various classes of asynchronous, distributed
optimization systems, demonstrating the use of the framework dis-
cussed in Part I [1]. The design and use of each class of systems is
based upon the following strategy:

1) Write a reduced-form optimization problem, defined in [1].
2) Connect appropriate constitutive relations to interconnection
elements, e.g. from Figs. 2-3 in [1], implementing the associ-
tated transformed stationarity conditions. Delay-free loops will
generally result.
3) Break delay-free loops:
   a) For any constitutive relation that is a source element,
      perform algebraic simplification thereby incorporating the
      solution of the algebraic loop into the interconnection.
   b) Insert synchronous or asynchronous delays between the
      remaining constitutive relations and the interconnection.
4) Run the distributed system until it reaches a fixed point.
The discussion in Section III, in conjunction with the system
properties in Fig. 3 in [1], provide guidance in determining when
convergence is ensured.
5) Write out the primal and dual decision variables \( a_i \) and \( b_i \)
   by multiplying the variables \( c_i \) and \( d_i \) by the inverses of the \((2 \times 2)\)
   matrices used in transforming the stationarity conditions.

II. EXAMPLE SYSTEMS

Figs. 2-7 depict various asynchronous, distributed optimization
algorithms implemented using the presented framework, specifically
making use of the elements in Figs. 2-3 of Part I [1]. Figs. 2 and 3
in this paper illustrate two alternative implementations of systems for
solving the LASSO problem. Figs. 4 and 5 depict two alternative
implementations of systems for performing minimax-optimal FIR
filter design. Fig. 6 depicts a support vector machine classifier
trained using a centralized algorithm generated using the presented
framework. Fig. 7 illustrates an example of a nonconvex optimization
algorithm aimed at the problem discussed in [2], in particular that of
designing a sparse FIR filter for acoustic equalization. In Figs. 2-
7, the asynchronous delay elements were numerically simulated using
discrete-time sample-and-hold systems triggered by independent
Bernoulli processes, with the probability of sampling being 0.1.

III. DISCUSSION OF CONVERGENCE

Fig. 1(a) summarizes the overall interconnection of elements
composing the presented class of systems discussed in Part I [1],
with those maps \( m(x) \) corresponding to source relationships being
written separately. Figs. 1(b)-(d) illustrate a set of manipulations
useful in analyzing convergence, with Fig. 1(b) specifically depicting
a solution to the transformed stationarity conditions. The approach is
to begin with the system in Fig. 1(a) and perform the additions and
subtractions of \( c_i^* \) and \( d_i^* \) indicated in Fig. 1(c), obtaining Fig. 1(d)
by identifying that Fig. 1(c) is a superposition of Figs. 1(b) and (d).

There are various ways that the system in Fig. 1(d) can be used
in determining necessary conditions for convergence, a subset of
which we outline here. Generally, arguments for convergence utilizing
Fig. 1(d) involve identifying conditions for which \( \|d^*_D\| \) in this figure
is strictly less than \( \|c^*_n\| \), except at 0. Using the definition of a source
element in [1] and the fact that \( G \) is a neutral map, i.e. an orthonormal
matrix, we conclude from Fig. 1(d) that

\[
\|d^*_D\| \leq \|c^*_n\|. \tag{1}
\]

If, for example, the solution to the transformed stationarity condi-
tions \( c_i^* \) and \( d_i^* \) is known to be unique, and additionally if the col-
lection of constitutive relations denoted \( m(\cdot) \) is known to be dissipative
about \( d^*_n \), then from Fig. 1(d) we conclude that \( \|c^*_n\| \leq \|d^*_n\| \)
except at 0, resulting in

\[
\|d^*_D\| < \|d^*_n\|. \tag{2}
\]

The authors wish to thank Analog Devices, Bose Corporation, and Texas
Instruments for their support of innovative research at MIT and within the
Digital Signal Processing Group.
Optimisation Problem
\[
\min \frac{1}{2} \| x \|^2 + \frac{\lambda}{2} \| a \|^2
\]
\begin{align*}
\text{s.t.} \quad & a = Ax - b \\
\frac{w}{\lambda} = \frac{c}{\lambda} \\
\gamma = \frac{y}{\lambda}
\end{align*}
P is a 1-norm approximation

\[
\widehat{Q}_k(a) = 0
\]
for each \( a \) in \( \mathbb{R} \)

\[
\lambda \geq \frac{1}{\lambda}
\]
\[
\lambda < \frac{1}{\lambda}
\]

except at \( 0 \). Eq. 2 implies, for example, that coupling the constitutive relations denoted \( m(\cdot) \) to the linear interconnection elements via deterministic vector delays, the discrete-time signal denoted \( d_m[n] \) will converge to 0 and so the signal \( d_\lambda[n] \) will converge to \( d_m \).

The uniqueness of the stationarity conditions and the property of the constitutive relations being dissipative used in the preceding argument are not, however, strictly required. A more general line of reasoning involves justifying Eq. 2 in the vicinity of any such solution \( c^* \) and \( d^* \), for example by claiming that even if specific constitutive relations \( m_k(\cdot) \) are norm-increasing, the overall interconnected system results in a map from \( d_m^* \) to \( d_\lambda^* \) that is norm-reducing in the vicinity of that solution.

Arguments for convergence involving essentially Eq. 2 can also be applied in a straightforward way to systems utilizing asynchronous delays, modeled as discrete-time sample-and-hold systems triggered by independent Bernoulli processes. In particular by taking the expected value of \( \| d_m[n] \|^2 \), applying the law of total expectation, substituting in Eq. 2, and performing algebraic manipulations, it can be argued that \( E[\| d_m[n] \|^2] \) converges to 0. A more formal treatment of convergence is the subject of future work.
Fig. 4. Signal processing architecture and numerical simulation corresponding to a minimax-optimal FIR filter design problem, specifically that of lowpass filter design. The obtained result is compared with a known solution from the Parks-McClellan algorithm.

Fig. 5. Alternative algorithm for minimax-optimal filter design, obtained by modification of the problem statement in Fig. 4 and intended to demonstrate that the presented framework can be used in designing a variety of distinct classes of algorithms. The parameter $\rho$ is selected to specify the relative enforcement of equality between the system variables loosely shared between the two linear interconnection elements. For the depicted solution $\rho$ is selected to be small, resulting in a very close approximation to the lowpass filter design problem in Fig. 4.
Fig. 6. Signal processing architecture for a single agent in a connected graph implementing a decentralized algorithm for training a support vector machine classifier. The numerical simulation depicts a system involving 30 such agents, each having knowledge of a single training vector. The parameter $\rho$ specifies the relative enforcement of equality for the system variables that are coupled between each agent in the graph. For the depicted solution $\rho$ is selected to be small, and the graph is known to be connected, with each node as depicted above having exactly four incident connections.

Fig. 7. Signal processing architecture and numerical simulation corresponding to a nonconvex sparse filter design problem. The parameters $\rho$ and $\nu$ are respectively selected to specify the enforcement of the size of $x$ and the width of the abrupt decrease in cost about $0$ for the nonconvex element. $\rho_+$ and $\rho_-$ affect the enforcement of the soft inequality constraints. For the depicted solution $\rho$ and $\rho_+$ are selected to be small and $\rho_-$ and $\nu$ are selected to be large.
REFERENCES


