A cyclic block coordinate gradient projection algorithm

Silvia Bonettini, Dipartimento di Matematica, Università di Ferrara

Motivation: Many real-life problems can be modeled as smooth, large-scale optimization problems whose variables are constrained in a cartesian product of convex sets.

\[
\begin{align*}
\min & \quad f(x) \\
n & \quad x \in \Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_m \subseteq \mathbb{R}^n \\
x & = (x_1, x_2, \ldots, x_m) \quad x_i \in \Omega_i
\end{align*}
\]

Basic idea: Decouple the optimization problem by cycling over the blocks

Nonlinear Gauss-Seidel method

\[
x_i^{(k+1)} = \arg \min_{y \in \Omega_i} f(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, y, x_{i+1}^{(k)}, \ldots, x_m^{(k)})
\]

Drawbacks of the nonlinear GS method:
- No convergence for \( m > 2 \) without strict convexity assumptions;
- Exact solution of a constrained minimization problem required at each partial iteration.

Recent developments:
- Convergence result for \( m = 2 \) in the nonconvex, constrained case [5];
- Globally convergent line-search based schemes for the unconstrained case [4].

Proposed approach: a descent method based on the projected gradient properties.

The Cyclic Block Gradient Projection (CBGP) method

Choose the starting point \( x^{(0)} \in \Omega \) and a positive integer \( L \)

For \( k = 0, 1, 2, \ldots \)

For \( i = 1, \ldots, m \)
- Choose the inner iteration number \( \ell_i^{(k)} \leq L \)
- Apply \( \ell_i^{(k)} \) SGP iterations to the problem

\[
\min_{y \in \Omega_i} J_i(y) = f(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, y, x_{i+1}^{(k)}, \ldots, x_m^{(k)})
\]

End

End

Convergence analysis: [1]

Let \( \{x^{(k)}\} \) the sequence generated by the CBGP algorithm and assume that \( x^* \) is a limit point of \( \{x^{(k)}\} \). Then, \( x^* \) is a stationary point.

The Scaled Gradient Projection (SGP) method

Choose the starting point \( y^{(0)} \in Y \) and set the parameters

\[
\beta \in (0, 1), 0 < \alpha_{\text{min}} < \alpha_{\text{max}}
\]

For \( \ell = 0, 1, 2, \ldots \)

- Choose the step-length \( \alpha \in [\alpha_{\text{min}}, \alpha_{\text{max}}] \) and a positive definite scaling matrix \( D \)
- Compute the descent direction

\[
d^{(\ell)} = P_Y D^{-1}(y^{(\ell)} - \alpha D y J_i(y^{(\ell)})) - y^{(\ell)}
\]

- Armijo Backtracking loop: compute \( \lambda_{\ell} \) such that

\[
J(y^{(\ell)} + \lambda_{\ell} d^{(\ell)}) \leq J(y^{(\ell)}) + \beta \lambda_{\ell} \nabla J_i(y^{(\ell)})^T d^{(\ell)}
\]

- Set \( y^{(\ell+1)} = y^{(\ell)} + \lambda_{\ell} d^{(\ell)} \)

End

Remarks: No convexity assumptions, no limitations on the numbers of blocks, convergence for any choice of inner iterations number (just bounded).

Implementation issues: exploit scaling matrix and step-length choices to improve convergence (e.g. Barzilai-Borwein step-length selection rules).

Applications: [1],[2],[3]

Nonnegative Matrix Factorization (NMF)

Given a data matrix \( V \in \mathbb{R}^{m \times n} \) and a positive integer \( r < m \), find \( W \in \mathbb{R}^{m \times r}, H \in \mathbb{R}^{r \times n} \) s.t.

\[
\min_{W \geq 0, H \geq 0} \frac{1}{2} \|WH - V\|_F^2
\]

Finds sparse representation of data

Semi-blind deconvolution from sparse Fourier data

Given the data \( g \in \mathbb{C}^n \), find the image \( f \in \mathbb{C}^n \), the frequencies \( \omega \in \mathbb{R}^{n+2} \) s.t.

\[
\min_{f \geq 0} \frac{1}{2} \| \omega(f) \|_2^2
\]

where \( \omega(f) = \sum_{j=1}^{n} \hat{f}_j e^{-2\pi i j \omega} \)

Applications to solar flares image reconstruction [2]

Blind deconvolution

Given the observed image \( g \in \mathbb{R}^{n \times n} \), find the true image \( f \in \mathbb{R}^{n \times n} \) and the psf \( h \in \mathbb{R}^{n \times n} \) s.t.

\[
\min_{f \geq 0} \quad \text{dist}(f \ast h, g)
\]

where \( \text{dist}(f \ast h, g) = \sum_{i,j} |f_{ij} - g_{ij}| \mid h_{ij} \mid \)

References: