# Primal-Dual Method for Huge-Scale Nonsmooth Optimization Problem

Sergey Shpirko, PreMoLab (MFTI), Russia

June 22, 2012

Joint paper with Yu.Nesterov (University of Louvain)

# Outline

# Updating the sparse matrix-vector product

**Notation:** Let $x = (x^{(1)}, \ldots, x^{(n)}) \in R^n$, $p(x) \leq k$ - number of nonzero elements, $\gamma(x) = \frac{p(x)}{dim\ x} \leq 1$ sparsity coefficient of vector x.

**Matrix-vector product**

General dense matrix $A \in R^{NM}$, an arbitrary $x \in R^n \Longrightarrow MN$ a.o.;

Sparse matrix $\Longrightarrow \gamma(A)MN$ a.o.;

**Recursive update of the product:**

Let us assume $y = Ax$ is already computed.

We need to compute a new vector

$$y_+ = Ax_+, \quad \text{where} \quad x_+ = x + d.$$

# Updating the sparse matrix-vector product

The complexity of this update is

$$\kappa_A(d) = \sum_{j \in \sigma(d)} p(Ae_j).$$

## Lemma

*Assume that the matrix $A \in R^{NM}$ has a uniform filling:*

$$\gamma(Ae_j) \leq c_u \gamma(A) \quad i = 1, ..., N,$$

*where $c_u \geq 1$ is an absolute constant.*
*Assume also that $\gamma(d) \leq c_u \gamma(A)$.*
*Then*

$$\kappa_A(d) \leq c_u^2 \gamma(A)^2 MN.$$

# Fast updates in short computational tree

$f(x^1, ..., x^n)$ **short-tree representable** if its value can be computed by a short binary tree with height $\ln n$.

*Level zero*:

$$v_{0,i} = x^i, \quad i = 1, ..., n;$$

*Level* $i$:

$$v_{i+1,j} = \psi_{i+1,j}(v_{i,2j-1}, v_{i,2j}), \quad j = 1, ..., 2^{k-i-1}, \quad i = 0, ..., k-1,$$

where $\psi_{i,j}$ are some functions of two variables.

**Examples of functions:**

$$f(x) = \|x\|_p, \quad p \geq 1, \qquad \psi_{i,j}(t_1, t_2) \equiv [|t_1|^p + |t_2|^p]^{1/p}$$
$$f(x) = \ln\left(\sum_{i=1}^n e^{x^i}\right), \qquad \psi_{i,j}(t_1, t_2) \equiv \ln(e^{t_1} + e^{t_2}),$$
$$f(x) = \max\{x^1, ..., x^n\}, \qquad \psi_{i,j}(t_1, t_2) \equiv \max\{t_1, t_2\}.$$

**Advantage:**

The computational $f(x_+)$ needs only $k \equiv \log_2 n$ applications of $\psi_{i,j}(\cdot, \cdot)$.

# Simple subgradient method

Consider an optimization problem with single functional constraint

$$\min_{x \in Q} \{ f(x) : g(x) \leq 0 \},$$

with $Q \in R^n$ - closed convex set, $f, g$ - convex functions, which subgradients are uniformly bounded on $Q$.

**Method** $SG_N(h)$

For $k = 0, .., N-1$ iterate:

$$\text{If } g(x_k) > h\|g'(x_k)\|, \quad \text{then (A) } x_{k+1} = \pi_Q \left( x_k - \frac{g(x_k)}{\|g'(x_k)\|^2} g'(x_k) \right),$$
$$\text{else (B) } x_{k+1} = \pi_Q \left( x_k - \frac{h}{\|f'(x_k)\|} f'(x_k) \right).$$

# Simple subgradient method

Consider an optimization problem with single functional constraint

$$\min_{x \in Q} \{ f(x) : g(x) \leq 0 \},$$

with $Q \in R^n$ - closed convex set, $f, g$ - convex functions, which subgradients are uniformly bounded on $Q$.

**Method** $SG_N(h)$

For $k = 0, .., N-1$ iterate:

$$\text{If } g(x_k) > h\|g'(x_k)\|, \quad \text{then (A) } x_{k+1} = \pi_Q \left( x_k - \frac{g(x_k)}{\|g'(x_k)\|^2} g'(x_k) \right),$$
$$\text{else (B) } x_{k+1} = \pi_Q \left( x_k - \frac{h}{\|f'(x_k)\|} f'(x_k) \right).$$

# Simple subgradient method

**Terminology:** $\mathcal{F}_k \subseteq \{0, \ldots, k\}$ - set of iteration type (B)

$$f_k^* = \min_{i \in \mathcal{F}_k} f(x_i), \quad L_k(f) = \max_{i \in \mathcal{F}_k} \|f'(x_i)\|,$$
$$g_k^* = \max_{i \in \mathcal{F}_k} g(x_i), \quad L_k(g) = \max_{i \in \mathcal{F}_k} \|g'(x_i)\|.$$

Lets fix some an arbitrary feasible point and note $r_k(x) \equiv \|x_k - x\|$.

**Theorem**

*If $N > r_0^2/h^2$, then $\mathcal{F}_k \neq \varnothing$ and*

$$f_N^* - f(x) \leq hL_N(f), \quad g_N^* \leq hL_N(g).$$

# Solving the huge-scale optimization problems

We assume that $A \in R^{M \times N}$ has *a block structure*: divided on $m \times n$ blocks $A_{i,j} \in R^{r_i \times q_j} : \sum_{i=1}^{m} r_i = M, \ \sum_{j=1}^{n} q_j = N$.
We assume that block row $A_i = (A_{i,1}, ..., A_{i,n})$ and block column $A^j = (A_{1,j}, ..., A_{m,j})$ is block-sparse:

$$p(A_{i,j}) = q_j r_i, \quad j \in \sigma_b(A_i) \subset \{1, ..., n\}, \quad i \in \sigma_b(A^j) \subset \{1, ..., m\}.$$

## Consider an optimization problem:

$$\min\{f(x) = f_0(u^0(x)) : \psi(u) = \max_{1 \le i \le m} f_i(u^i), \ g(x) = \psi(u(x)) \le 0,$$

$$u^i(x) = \sum_{j \in \sigma_b(A_i)} A_{i,j} x^j - b^j, \ i = 1, ..., m, \quad x^j \in Q_j, \ j = 1, ..., n\},$$

where $f_i(u^i)$ , $i = 0, ..., m$ - convex functions and have bounded subgradients, $Q_j$ - convex and closed.

# Solving the huge-scale optimization problems

We assume that $A \in R^{M \times N}$ has *a block structure*: divided on $m \times n$ blocks $A_{i,j} \in R^{r_i \times q_j} : \sum_{i=1}^{m} r_i = M, \ \sum_{j=1}^{n} q_j = N$.
We assume that block row $A_i = (A_{i,1}, ..., A_{i,n})$ and block column $A^j = (A_{1,j}, ..., A_{m,j})$ is block-sparse:

$$p(A_{i,j}) = q_j r_i, \quad j \in \sigma_b(A_i) \subset \{1, ..., n\}, \quad i \in \sigma_b(A^j) \subset \{1, ..., m\}.$$

## Consider an optimization problem:

$$\min\{f(x) = f_0(u^0(x)) : \psi(u) = \max_{1 \le i \le m} f_i(u^i), \ g(x) = \psi(u(x)) \le 0,$$

$$u^i(x) = \sum_{j \in \sigma_b(A_i)} A_{i,j} x^j - b^j, \ i = 1, ..., m, \quad x^j \in Q_j, \ j = 1, ..., n\},$$

where $f_i(u^i)$ , $i = 0, ..., m$ - convex functions and have bounded subgradients, $Q_j$ - convex and closed.

# Solving the huge-scale optimization problems

Let $\delta_k^j = x_{k+1}^j - x_k^j$ and $x_{k+1} = \pi_Q(x_k + d_k)$. Assume that $d_k$ - block-sparsed. Then

$$x_{k+1}^j = \pi_{Q_j}(x_k^j + d_k^j), \quad j \in \sigma_b(d_k) \longrightarrow c_\pi p_b(d_k) \quad \text{a.o.}$$

$$x_{k+1}^j = x_k^j, \quad \textit{otherwise}$$

Assume that $u_k = Ax_k - b$ is already computed. Then residual $u_{k+1} = Ax_{k+1} - b$ can be obtained by sequence of recursive updates:

$u_+ = u_k$; **For** $j \in \sigma_b(d_k)$, $i \in \sigma(A^j)$ **iterate:**

1. Update $u_+^i = u_+^i + A_{i,j}\delta_k^j \longrightarrow r_i q_j$ a.o.;
2. Compute $f_i(u_+^i)$, $f_i'(u_+^i) \longrightarrow c_f r_i$ a.o.;

3. Update $\psi(u_+)$, $i_+ = argmax_{1 \le i \le m} f_i(u_+^i) \longrightarrow \log_2 m$ a.o.;

$u_{k+1} = u_+$

# Solving the huge-scale optimization problems

**Theorem**

*Assume that filling of matrix $A$ is uniform:*

$$\frac{1}{r_i} p(A_i) \leq \frac{c_u}{M} p(A), \quad p_b(A_i) \leq \frac{c_u}{M} p_b(A), \quad i = 1, ..., m,$$

$$\frac{1}{q_j} p(A^j) \leq \frac{c_u}{N} p(A), \quad p_b(A^j) \leq \frac{c_u}{m} p_b(A), \quad j = 1, ..., n.$$

*Then the computational costs don't exeed*

$$c_u^2 [\gamma^2(A) MN + \gamma_b^2(A) mn \log_2 m].$$

# Primal-dual subgradient method for LP

Consider the following primal-dual pair of LP:

$$f^* = \min\{\langle c, x\rangle : Ax = b,\ x \geq 0\} =$$

$$= \max_{s \in R^n, y \in R^m} \{\langle b, y\rangle : s = c - A^T y \geq 0\}.$$

Here $c \in R^n,\ b \in R^m,\ A \in R^{m \times n}$.
Assume that both problems are solvable with no duality gap. Then

$$\exists x^* \geq 0,\ y^* \in R^M : Ax^* = b,\ s^* = c - A^T y^* \geq 0,\ \langle s^*, x^*\rangle = 0.$$

For $y \in R^m$ denote

$$j(y) : \frac{\langle Ae_{j(y)}, y\rangle - c^{j(y)}}{\|Ae_{j(y)}} = g(y) \equiv \max_{1 \leq j \leq n} \frac{\langle Ae_j, y\rangle - c^j}{\|Ae_j\|},$$

$$g'(y) = \frac{Ae_j(y)}{\|Ae_j\|},\ \|g'(y)\| = 1.$$

# Primal-dual subgradient method for LP

## How the method works

$$y_0 = 0; \quad \text{For} \quad k \geq 0 \quad \text{do:}$$

$$\text{If} \quad g(y_k) \leq h, \text{then (F):} \quad y_{k+1} = y_k + h\frac{b}{\|b\|},$$

$$\text{else (G):} \quad y_{k+1} = y_k - g(y_k)g'(y_k).$$

# Primal-dual subgradient method for LP

## How the method works

$$y_0 = 0; \quad \text{For} \quad k \geq 0 \quad \text{do:}$$

$$\text{If} \quad g(y_k) \leq h, \text{then (F):} \quad y_{k+1} = y_k + h\frac{b}{\|b\|},$$

$$\text{else (G):} \quad y_{k+1} = y_k - g(y_k)g'(y_k).$$

**Terminology:**

For $N > 0$ denote by $\mathcal{F}_N$ the set of iteration type (F) and $\mathcal{G}_N = \{0, ..., N\} \backslash \mathcal{F}_N$, $N_f = |\mathcal{F}_N|$, $N_g = |\mathcal{G}_N|$.

# Primal-dual subgradient method for LP

Let us define the approximations for primal and dual solutions:

$$\bar{x}_N = \frac{\|b\|}{hN_f} \sum_{k \in \mathcal{G}_N} \frac{g(y_k)}{\|Ae_{j(y_k)}\|} e_{j(y_k)}, \quad \bar{y}_N = \frac{1}{N_f} \sum_{k \in \mathcal{F}_N} y_k, \; \bar{s}_N = c - A^T \bar{y}_N.$$

**Motivation:**

$$\bar{s}_N = c - \frac{1}{N_f} \sum_{k \in \mathcal{F}} A^T \bar{y}_k \geq -hd_A, \text{ where } d_A \in R^n : d_A^{(j)} = \|Ae_j\|,$$

$$y_{N+1} = \frac{hN_f}{\|b\|} b - \sum_{k \in \mathcal{G}_N} \frac{g(y_k)}{\|Ae_{j(y_k)}\|} Ae_{j(y_k)} =$$

$$= \frac{hN_f}{\|b\|} (b - A\bar{x}_N).$$

# Primal-dual subgradient method for LP

## Theorem

*Denote* $D = 2\left(\frac{\langle x^*, d_A\rangle}{\|b\|} + 1\right)$. *For any* $N \geq 0$ *we have:*

$$N_f \geq \frac{1}{D}\left(N + 1 - \frac{\|y^*\|^2}{h^2}\right).$$

*If* $N_f \geq 1$, *then* $\langle c, \bar{x}_N\rangle - \langle b, \bar{y}_N\rangle \leq \frac{1}{2}h\|b\|$. *Finally, if*

$$N + 1 > \frac{\|y^*\|^2}{h^2}, \tag{1}$$

*then* $\langle x^*, s\bar{s}_N\rangle + \langle \bar{x}_N, s^*\rangle \leq h\|b\|$,
*and the residual in primal-dual system vanishes as* $N \to \infty$*:*

$$\frac{1}{\|b\|}\|b - A\bar{x}_N\| \leq \sqrt{\frac{D}{N_f}} + \frac{\|y^*\|}{hN_f}.$$

# Primal-dual subgradient method for LP

## PDM with explicit rules for the choice of parameters

We have three positive accuracy parameters $\epsilon_f,\ \epsilon_g,\ \epsilon_a$. Its goal to generate an approximate primal-dual solution $(\bar{x}_N,\ \bar{y}_N,\ \bar{s}_N)$:

$$\bar{x}_N \geq 0, \quad \bar{s}_N = c - A^T \bar{y}_N \geq -\epsilon_g, \langle c, \bar{x}_N \rangle - \langle b, \bar{y}_N \rangle \leq \epsilon_f, \qquad (2)$$

$$\text{with following stopping criterion} \quad \|A\bar{x}_N - b\| \leq \ \epsilon_a. \qquad (3)$$

## Theorem

*Let* $h = \min \left\{ \epsilon_f \cdot \frac{2}{\|b\|},\ \epsilon_g \cdot \frac{1}{\max\limits_{1 \leq j \leq n} \|Ae_j\|} \right\}$

*Then inequalities (2) are satisfied automatically.*

*The complexity bound for reaching (3)*

$$N + 1 \geq \left(\frac{r_0}{h}\right)^2 + D \cdot \max \left\{ \frac{2r_0\|b\|}{h\epsilon_a}, \frac{4\|b\|^2 D}{\epsilon_a^2} \right\}.$$