Corrections and Notes for

CONVEX OPTIMIZATION ALGORITHMS

by Dimitri P. Bertsekas

Athena Scientific

Last Updated: 11/1/17

p. 11 (+4) Change "$x_2 \in \Re^n$" to "$x_2 \in \Re^m$"

p. 11 (-12) Change "at f at x" to "of f at x"

p. 11 (-3) Change "$z \in \Re^n$" to "$z \in \Re^m$"

p. 15 (-5) Change "S" to "$S^\perp$"

p. 46 (+5) Change the statement and proof of Prop. 1.5.3 as follows:

Proposition 1.5.3: Let $f : Y \mapsto \Re$ be a function defined on a subset Y of $\Re^n$, and let X_i, $i = 1, \ldots, m$, be closed subsets of Y with nonempty intersection. Assume that f is Lipschitz continuous over Y with constant L, and that for some scalar $\beta > 0$, we have

$$\mathrm{dist}(x; X_1 \cap \cdots \cap X_m) \leq \beta \sum_{i=1}^{m} \mathrm{dist}(x; X_i), \qquad \forall\, x \in Y.$$

Let c be a scalar such that $c > \beta L$. Then the set of minima of f over $\cap_{i=1}^m X_i$ coincides with the set of minima of

$$f(x) + c \sum_{i=1}^{m} \mathrm{dist}(x; X_i)$$

over Y.

Proof: The proof is similar to the proof of Prop. 1.5.2, using the given additional condition to modify the main inequality. Denote $F(x) = f(x) + c \sum_{i=1}^m \mathrm{dist}(x; X_i)$ and $X = X_1 \cap \cdots \cap X_m$. For a vector $x \in Y$, let $\hat{x}_i$

denote a vector of X_i that is at minimum distance from x, and let $\hat{x}$ denote a vector of X that is at minimum distance from x. If $c > \beta L$, we have for all $x \in Y$,

$$\begin{aligned} F(x) &= f(x) + c \sum_{i=1}^{m} \|x - \hat{x}_i\| \\ &\geq f(\hat{x}) + \big(f(x) - f(\hat{x})\big) + \frac{c}{\beta} \|x - \hat{x}\| \\ &\geq f(\hat{x}) + \left(\frac{c}{\beta} - L\right) \|x - \hat{x}\| \\ &\geq F(\hat{x}), \end{aligned}$$

with strict inequality if $x \neq \hat{x}$. The proof now proceeds as in the proof of Prop. 1.5.2. **Q.E.D.**

p. 92 (+8) For clarity change
"More specifically, instead of the gradient sum

$$s_k = \sum_{\ell=0}^{m-1} \nabla f_{i_{k-\ell}}(x_{k-\ell}),$$

in Eq. (2.35), these methods use

$$\tilde{s}_k = \nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}_k) + \sum_{\ell=0}^{m-1} \nabla f_{i_{k-\ell}}(\tilde{x}_k),$$

where $\tilde{x}_k$ is the most recent point where the full gradient has been calculated. To calculate $\tilde{s}_k$ one only needs to compute the difference of the two gradients

$$\nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}_k)$$

and add it to the full gradient $\sum_{\ell=0}^{m-1} \nabla f_{i_{k-\ell}}(\tilde{x}_k)$."
to
"For an example, instead of the gradient sum

$$s_k = \sum_{\ell=0}^{m-1} \nabla f_{i_{k-\ell}}(x_{k-\ell}),$$

in Eq. (2.35), such a method may use $\tilde{s}_k$, updated according to

$$\tilde{s}_k = \nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}_k) + \tilde{s}_{k-1},$$

where $\tilde{s}_0$ is the full gradient computed at the start of the current cycle, and $\tilde{x}_k$ is the point at which this full gradient has been calculated. Thus to obtain $\tilde{s}_k$ one only needs to compute the difference of the two gradients

$$\nabla f_{i_k}(x_k) - \nabla f_{i_k}(\tilde{x}_k)$$

and add it to the current approximation of the full gradient $\tilde{s}_{k-1}$."

p. 113 (-13) Change "Lagrangian" to "Lagrangian method"

p. 206 (+13) Change "$\tilde{\lambda} \in \partial f(x_{\tilde{\lambda}})$" to "$\tilde{\lambda} \in \partial f(x)$"

p. 255 (-5) Change "$-\|\overline{x}_1 - \overline{x}_2\|^2$" to "$-\frac{1}{c}\|\overline{x}_1 - \overline{x}_2\|^2$"

p. 283 (+15) Change "(5.83)" to "(5.87)"

p. 283 (+18) Change "$Ax = z$" to "$Ax + Bz = d$"

p. 284 (+2) Change "λ^1_{k+1}" and "λ^m_{k+1}" to "λ^1_k" and "λ^m_k", respectively

p. 293 (+2) Change "Example 3.4" to "Section 3.4, Example 3.4"

p. 293 (+4) Change Eq. (5.107) and the following line to

$$\lambda^j_{k+1} = \lambda^j_k + \frac{c}{m_j}\left(\sum_{i=1}^{m} A_{ji}x^i_{k+1} - b^j\right), \qquad j = 1, \ldots, r, \tag{5.107}$$

where A_{ji}, b^j, and r are the components and row dimension of A_i and b, respectively, and m_j is the number of

p. 298 (+8) Change "$z \in \Re^n$" to "$z \in \Re^m$"

p. 299 (-5) Change "$\lambda - cv$" to "$\lambda + cv$"

p. 300 (-6) Change "f_1 (or f_2, respectively)" to "f_2 (or f_1, respectively)"

p. 308 (-9) Change "$y \in \Re^n$" to "$y \in X$"

p. 314 (+9 to +22) Replace the part of the proof starting just after the equation $\phi(x) = f(x) - \frac{\sigma}{2}\|x\|^2$. and ending just before the statement "Using the expression (6.25) ..." with the following:

We will show that $\nabla\phi$, which is given by

$$\nabla\phi(x) = \nabla f(x) - \sigma x, \tag{6.25}$$

is Lipschitz continuous with constant $L - \sigma$. To this end, based on the equivalence of statements (i) and (v) of Exercise 6.1, it is sufficient to show that

$$\big(\nabla\phi(x) - \nabla\phi(y)\big)'(x - y) \le (L - \sigma)\|x - y\|^2, \qquad \forall\ x, y \in \Re^n,$$

or, using the expression (6.25) for $\nabla\phi$,

$$\big(\nabla f(x) - \nabla f(y) - \sigma(x - y)\big)'(x - y) \le (L - \sigma)\|x - y\|^2, \qquad \forall\ x, y \in \Re^n.$$

This relation is equivalently written as

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) \le L\|x - y\|^2, \qquad \forall\ x, y \in \Re^n,$$

and is true by (iii) of part (a).

Having shown that $\nabla\phi$ is Lipschitz continuous with constant $L-\sigma$, we apply (ii) of part (a) to the function ϕ and obtain

$$\big(\nabla\phi(x)-\nabla\phi(y)\big)'(x-y) \geq \frac{1}{L-\sigma}\big\|\nabla\phi(x)-\nabla\phi(y)\big\|^2.$$

p. 316 (-17) Change "$\nabla f(\overline{x})=0$" to "the optimality condition $\nabla f(\overline{x})'(x-\overline{x}) \geq 0$, for all $x \in X$"

p. 318 (-13) Change "β^m" to "$\sigma\beta^m$"

p. 323 (+7) Change "gradient [cf. Eq. (6.6)]" to "gradient within $\Re^n$"

p. 323 (+12) Change "$\beta_k \in (0,1)$" to "$\beta_k \in [0,1)$"

p. 323 (-11) Change "$\theta_0=\theta_1 \in (0,1]$" to "$\theta_{-1}=\theta_0=1$"

p. 323 (-10) In Eq. (6.33) add the additional condition $\theta_k \leq 2/(k+2)$, which is used in the proof of Prop. 6.2.1 [p. 326, (-16)]. So Eq. (6.33) should now read

$$\frac{1-\theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2}, \qquad \theta_k \leq \frac{2}{k+2}, \qquad k=0,1,\ldots. \tag{6.33}$$

p. 323 (-1) Change "$x_{k+1}=P_X\big(x_k-\alpha\nabla f(x_k)\big)$," to "$x_{k+1}=P_X\big(y_k-\alpha\nabla f(y_k)\big)$,"

p. 324 (-8) Change "∇f satisfies the Lipschitz condition (6.6)" to "∇f is Lipschitz continuous within $\Re^n$ with Lipschitz constant L,"

p. 336 (-16) Change "$x \in \Re^n$" to "$x \in \Re^m$"

p. 365 (-5) Change the statement of Prop. 6.4.9 in accordance with the correction to Prop. 1.5.3.

p. 383 (+10) Change "whwew" to "where"

p. 428 (-19) Change "3.3" to "Section 3.3"

p. 437 (+13) Change "Note:" to "Notes: If Eq. (6.238) holds with $\gamma=1$ it holds for all $\gamma \in (1,2)$, so this exercise shows arbitrarily fast superlinear convergence for the case of a sharp minimum, even when X^* contains multiple points (cf. Exercise 6.4)."

p. 461 (+7) Change "$\beta_k \geq 0$, $\gamma_k > 0$" to "$0 \leq \beta_k$, $0 < \gamma_k \leq 1$"

p. 471 (-12) Change "dom(f)" to "$\Re^n$"

p. 485 (-7,-8) Erase these two lines

p. 488 (+8) Change "By the Conjugacy Theorem [Prop. 1.6.1(c)] C^* is equal to $\mathrm{cl}\,\delta_C$. Thus the polar cone of C^* is $\mathrm{cl}(C)$." to "By the Conjugacy Theorem [Prop. 1.6.1(d)], the polar cone of C^* is $\mathrm{cl}(C)$."

p. 511 (+3) Change "$z \in \Re^n$" to "$z \in \Re^m$"