Spring 2025

Lecture II: Proximal-point (PP) and accelerated PP algorithms for convex optimization

UFSC

Last revised in: xx/xx/xxxx

In this lecture, we will study the proximal-point algorithm for convex optimization and its iteration-complexity analysis.

1 The proximal-point algorithm for convex optimization

The proximal-point (PP) algorithm is a conceptual rather than a practical method. It is the baseline algorithm for the development of other important computational schemes in numerical optimization, including the proximal gradient method.

Consider the *convex optimization problem*

$$\boxed{\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x)} \tag{1}$$

where $f\colon \mathbb{R}^n \to (-\infty, \infty]$ is a proper convex and lower semicontinuous (lsc) function.

We also denote the *optimal value* of f by

$$\boxed{f_{opt} = \inf_{x\in\mathbb{R}^n} f(x)} \tag{2}$$

and assume that $f_{opt} > -\infty$. Some results of convergence rates can be obtained without assuming that the solution set of (1) is nonempty (see, e.g., [3]).

1.1 The exact PP algorithm

The main idea behind the method is to replace (1) by a "regularized version" based on the following operation

$$\boxed{\min_{x\in\mathbb{R}^n}\left\{f(x) + \frac{1}{2\lambda}\|x-z\|^2\right\}} \tag{3}$$

where $\lambda > 0$ is regularization parameter and $z \in \mathbb{R}^n$. Problem (3) has a unique solution, say $z_\lambda \in \mathbb{R}^n$, and $z_\lambda = z$ (i.e., it is a fixed point of the operation described in (3)) if and only if z is a solution of the minimization problem (1).

Motivated by this, the PP algorithm is defined by iterating the operation (3):

The (exact) PP method for solving (1) is defined as:

Let $x_0 \in \mathbb{R}^n$ and iterate for $k \geq 1$:

$$\boxed{x_k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2\lambda_k} \|x - x_{k-1}\|^2 \right\}} \tag{4}$$

where $\lambda_k > 0$.

Example 1.1. *As we already mentioned, the* PP algorithm *is a conceptual method; its study provides insight for the development and analysis of different practical algorithms for convex optimization. However, in order to illustrate the iteration mechanism of the method, we will apply it to solve a (convex) quadratic programming problem:*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \left\{ f(x) := \frac{1}{2}\langle x, Ax \rangle - \langle b, x \rangle \right\} \tag{5}$$

where $A \colon \mathbb{R}^n \to \mathbb{R}^n$ is a self-adjoint positive bounded linear operator and b is a vector in $\mathbb{R}^n$. Recall that Problem (5) is equivalent to the linear operator equation

$$Ax = b. \tag{6}$$

The PP method *(4) applied to f as in (5) reads as, for all $k \geq 1$,*

$$x_k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2}\langle x, Ax \rangle - \langle b, x \rangle + \frac{1}{2\lambda_k} \|x - x_{k-1}\|^2 \right\}$$

which is to say that $x = x_k$ is the unique solution of the (positive definite) regularized linear equation

$$\left(A + \frac{1}{\lambda_k} I \right) x = b + \frac{1}{\lambda_k} x_{k-1}. \tag{7}$$

Summarizing, the PP algorithm *for solving (5) or, equivalently, for solving the operator equation (6), proceeds by computing solution of regularized "approximations" of (6), namely (7).*

In the next subsection, we will analyze the convergence rates of the PP algorithm, as described in (4). It will be convenient to note that the iteration of the PP method can also be written as

$$\boxed{v_k := \frac{1}{\lambda_k} (x_{k-1} - x_k) \in \partial f(x_k)} \tag{8}$$

where $\partial f(x_k)$ denotes the subdifferential of f at x_k.

1.1.1 Convergence rates

We begin with the following proposition, by studying the progress of the iterations defined by the PP algorithm (4) both in terms of the "metric" $\|\cdot\|$ and the objective function f.

Proposition 1.2. *The following holds:*

(a) *For all* $x \in \mathbb{R}^n$ *and* $k \geq 1$,

$$\boxed{\|x_{k-1} - x\|^2 \geq \|x_k - x\|^2 + \|x_{k-1} - x_k\|^2 + 2\lambda_k\big(f(x_k) - f(x)\big).}$$

(b) *For all* $k \geq 1$,

$$\boxed{f(x_{k-1}) \geq f(x_k) + \langle v_k, x_{k-1} - x_k\rangle = f(x_k) + \lambda_k\|v_k\|^2.}$$

Proof. (a) Using the identity $\|a\|^2 - \|b^2\| = \|a-b\|^2 + 2\langle a-b, b\rangle$ with $a = x_{k-1} - x$ and $b = x_k - x$, we find

$$\|x_{k-1} - x\|^2 - \|x_k - x\|^2 = \|x_{k-1} - x_k\|^2 + 2\langle x_{k-1} - x_k, x_k - x\rangle. \tag{9}$$

On the other hand, direct use of (8) and the definition of $\partial f(x_k)$ yields

$$\langle v_k, x_k - x\rangle \geq f(x_k) - f(x), \tag{10}$$

which then combined with the definition of v_k as in (8) gives

$$\langle x_{k-1} - x_k, x_k - x\rangle \geq \lambda_k\big(f(x_k) - f(x)\big). \tag{11}$$

The desired inequality now follows from directly from (9) and (11).

(b) The inequality follows directly from (10) with $x = x_{k-1}$. The identity follows from the inequality combined with the definition of v_k as in (8). □

Remark 1.3. *Since the function* $f(\cdot) + \frac{1}{2\lambda_k}\|\cdot - x_{k-1}\|^2$ *is* $\frac{1}{\lambda_k}$*-strongly convex and, by the definition of* x_k*,*

$$0 \in \partial\left(f(\cdot) + \frac{1}{2\lambda_k}\|\cdot - x_{k-1}\|^2\right)(x_k),$$

it follows that

$$f(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 \geq f(x_k) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 + \frac{1}{2\lambda_k}\|x - x_k\|^2 \qquad \text{for all } x \in \mathbb{R}^n,$$

which clearly gives the inequality in Proposition 1.2(a).

The following lemma shows the monotonicity of the sequence of norms of v_k:

Lemma 1.4. *For all* $k \geq 1$*, we have*

$$\boxed{\|v_{k+1}\| \leq \|v_k\|.} \tag{12}$$

Proof. Direct use of (8) gives $v_k \in \partial f(x_k)$ and $v_{k+1} \in \partial f(x_{k+1})$. Using the fact that ∂f is monotone, we have

$$\langle v_{k+1} - v_k, x_{k+1} - x_k\rangle \geq 0.$$

Since $\lambda_{k+1}v_{k+1} = x_k - x_{k+1}$ (see (8)) and $\lambda_{k+1} > 0$ direct substitution yields $\langle v_{k+1} - v_k, v_{k+1}\rangle \leq 0$ and so

$$\|v_{k+1}\|^2 = \langle v_{k+1}, v_{k+1}\rangle \leq \langle v_k, v_{k+1}\rangle \leq \|v_k\|\|v_{k+1}\|,$$

which is clearly equivalent to the desired inequality. □

The convergence rates for the PP algorithm will follow from the above results combined with the following technical lemma:

Lemma 1.5. *Let (α_k), (λ_k), (β_k) and (γ_k) be sequences of nonnegative real numbers such that, for all $k \geq 1$,*

$$\begin{aligned} \alpha_{k-1} &\geq \alpha_k + \lambda_k \beta_k + \lambda_k^2 \gamma_k, \\ \beta_{k-1} &\geq \beta_k + \lambda_k \gamma_k. \end{aligned} \tag{13}$$

Define $\Delta_0 = 0$, $\Delta_k = \sum_{j=1}^k \lambda_j$ and $\widetilde{\Delta}_k = \sum_{j=1}^k \lambda_j \Delta_j$. Let also $\underline{\lambda} > 0$ be such that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then the following holds:

(a) $\alpha_{k-1} + \Lambda_{k-1}\beta_{k-1} \geq \alpha_k + \Lambda_k \beta_k + \lambda_k \Delta_k \gamma_k$.

(b) $\sum_{j=1}^k \lambda_j \Delta_j \gamma_j + \Delta_k \beta_k + \alpha_k \leq \alpha_0$.

(c) $\beta_k \leq \dfrac{\alpha_0}{\Delta_k} \leq \dfrac{\alpha_0}{\underline{\lambda} k}$.

(d) $\min\limits_{j=1,\ldots,k} \gamma_j \leq \dfrac{\alpha_0}{\widetilde{\Delta}_k} \leq \dfrac{2\alpha_0}{\underline{\lambda}^2 k^2}$.

(e) *If (γ_k) is nonincreasing, then the assertion in item* (d) *holds with γ_k instead of $\min_{j=1,\ldots,k} \gamma_j$.*

Proof. (a) Using (13), we find

$$\begin{aligned} \alpha_{k-1} + \Lambda_{k-1}\beta_{k-1} &\geq \left(\alpha_k + \lambda_k \beta_k + \lambda_k^2 \gamma_k\right) + \Lambda_{k-1}\beta_{k-1} \\ &\geq \left(\alpha_k + \lambda_k \beta_k + \lambda_k^2 \gamma_k\right) + \Lambda_{k-1}\left(\beta_k + \lambda_k \gamma_k\right) \\ &= \alpha_k + (\underbrace{\Delta_{k-1} + \lambda_k}_{=\Delta_k})\beta_k + \lambda_k(\underbrace{\Delta_{k-1} + \lambda_k}_{=\Delta_k})\gamma_k, \end{aligned}$$

which finishes the proof of (a).

(b) This follows from item (a) and a simple telescopic sum argument. (recall that $\Delta_0 = 0$.)

(c) This follows from item (b), the assumption that $\lambda_k \geq \underline{\lambda}$ for all $k \geq 1$ and the definition of Δ_k.

(d) The first inequality follows from item (b) and the definition of $\widetilde{\Delta}_k$. To prove the second inequality, note first that $\Delta_j = \sum_{\ell=1}^j \lambda_\ell \geq j\,\underline{\lambda}$. Hence,

$$\widetilde{\Delta}_k = \sum_{j=1}^k \lambda_j \Delta_j \geq \underline{\lambda} \sum_{j=1}^k \Delta_j \geq \underline{\lambda}^2 \sum_{j=1}^k j = \underline{\lambda}^2 \frac{k(k+1)}{2} \geq \underline{\lambda}^2 \frac{k^2}{2},$$

which in turn, when combined with the first inequality, finishes the proof of item (d).

(e) This follows trivially from the fact that, in this case, $\gamma_k = \min_{j=1,\ldots,k} \gamma_j$. □

Assumption 1.6. *Suppose Problem* (1) *admits at least one solution and denote by* $\operatorname{Zer} \partial f$ *the solution set of* (1).

Next we prove global convergence rates for the PP algorithm, both in terms of function values and for the "residual" of the inclusion (8):

Proposition 1.7. *Consider the sequence (x_k) evolved by the (exact)* PP algorithm (4) *and let (v_k) be as in* (8)*. Let d_0 denote the distance of x_0 to the solution set of $\operatorname{Zer}\partial f \neq \emptyset$ of* (1)*, i.e., $d_0 \coloneqq \min_{x\in\operatorname{Zer}\partial f}\|x-x_0\|$. Also, let Δ_k and $\widetilde{\Delta}_k$ be as in* Lemma 1.5*. Assume that there exists $\underline{\lambda}>0$ such that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then the following holds:*

(a) *For all $k \geq 1$,*

$$\boxed{f(x_k) - f_{opt} \leq \frac{d_0^2}{2\Lambda_k} \leq \frac{d_0^2}{2\underline{\lambda}\, k} = O\left(\frac{1}{k}\right).}$$

(b) *For all $k \geq 1$,*

$$\boxed{\|v_k\| \leq \frac{d_0}{\sqrt{\widetilde{\Delta}_k}} \leq \frac{\sqrt{2}d_0}{\underline{\lambda}\, k} = O\left(\frac{1}{k}\right).}$$

Proof. Let $\overline{x} \in \operatorname{Zer}\partial f$ be such that $d_0 = \|\overline{x} - x_0\|$. In particular, $f_{opt} = f(\overline{x})$.

Let

$$\alpha_k \coloneqq \frac{1}{2}\|x_k - \overline{x}\|^2, \quad \beta_k \coloneqq f(x_k) - f(\overline{x}) \quad \text{and} \quad \gamma_k \coloneqq \frac{1}{2}\|v_k\|^2.$$

By dividing the inequality in Proposition 1.2(a) (with $x = \overline{x}$) by $1/2$, using the above definitions and the definition of v_k as in (8) we get

$$\alpha_{k-1} \geq \alpha_k + \lambda_k\beta_k + \lambda_k^2\gamma_k.$$

On the other hand, using Proposition 1.2(b) and the definitions of α_k and γ_k we have

$$\alpha_{k-1} \geq \alpha_k + \lambda_k\gamma_k.$$

Hence, one can apply the results of the technical Lemma 1.5. Note the item (a) is a direct consequence of Lemma 1.5(c). On the other hand, item (b) follows from Lemma 1.5((d) and (e)) combined with Lemma 1.4. □

Remark 1.8. Proposition 1.7 *guarantees, in particular, that for a given tolerance $\rho > 0$, the* PP algorithm *finds $x \in \mathbb{R}^n$ (an approximate solution for* (1)*) satisfying*

$$f(x) - f_{opt} \leq \rho$$

in at most

$$\left\lceil \frac{d_0^2}{\underline{\lambda}\rho} \right\rceil$$

iterations, where $\lceil\theta\rceil \coloneqq \min\{n \in \mathbb{Z} \mid n \geq \theta\}$.

1.1.2 Convergence rates for level-bounded functions

In this subsection we present an alternative analysis of the convergence rates for the PP algorithm.

We start with some discrete inequalities. First recall that the largest root of the quadratic $a\theta^2 + b\theta - c$, where $a \neq 0$ and $c > 0$, is given by

$$\frac{2c}{b + \sqrt{b^2 + 4ac}}. \tag{14}$$

Some discrete inequalities.

Lemma 1.9 (Nesterov). *Assume that $\alpha_k \geq 0$ satisfies, for all $k \geq 1$,*

$$\boxed{\alpha_{k-1} \geq \alpha_k + \alpha_k^2\,.} \tag{15}$$

Then, for all $k \geq 1$,

$$\boxed{\alpha_k \leq \frac{\alpha_0}{1 + c\,\alpha_0\,k} \leq \frac{1}{c\,k} = O\left(\frac{1}{k}\right)} \tag{16}$$

where

$$c := \frac{2}{1 + \sqrt{1 + 4\alpha_0}}.$$

Proof. Assume that $\alpha_k > 0$ (if $\alpha_k = 0$ then the result is trivial). Using (15) and (14) we obtain

$$\alpha_k \leq \frac{2\alpha_{k-1}}{1 + \sqrt{1 + 4\alpha_{k-1}}}, \text{ and so } \frac{1}{\alpha_k} \geq \left(\frac{1 + \sqrt{1 + 4\alpha_{k-1}}}{2}\right)\frac{1}{\alpha_{k-1}}.$$

Note that

$$\frac{1 + \sqrt{1 + 4\alpha_{k-1}}}{2} = 1 + \delta \iff 2\delta = \sqrt{1 + 4\alpha_{k-1}} - 1 \iff \delta = \frac{2\alpha_{k-1}}{1 + \sqrt{1 + 4\alpha_{k-1}}},$$

that is,

$$\frac{1 + \sqrt{1 + 4\alpha_{k-1}}}{2} = 1 + \frac{2\alpha_{k-1}}{1 + \sqrt{1 + 4\alpha_{k-1}}}.$$

Hence,

$$\frac{1}{\alpha_k} \geq \left(1 + \frac{2\alpha_{k-1}}{1 + \sqrt{1 + 4\alpha_{k-1}}}\right)\frac{1}{\alpha_{k-1}} = \frac{1}{\alpha_{k-1}} + \frac{2}{1 + \sqrt{1 + 4\alpha_{k-1}}}.$$

In view of (15) we have, in particular, $\alpha_0 \geq \alpha_{k-1}$ and so

$$\begin{aligned}\frac{1}{\alpha_k} &\geq \frac{1}{\alpha_{k-1}} + \frac{2}{1 + \sqrt{1 + 4\alpha_{k-1}}} \\ &\geq \frac{1}{\alpha_{k-1}} + \frac{2}{1 + \sqrt{1 + 4\alpha_0}} \\ &= \frac{1}{\alpha_{k-1}} + c.\end{aligned}$$

Consequently,

$$\frac{1}{\alpha_k} \geq \frac{1}{\alpha_0} + c\,k = \frac{1 + c\,\alpha_0 k}{\alpha_0},$$

which gives the desired result. □

Corollary 1.10 (Generalization of Lemma 1.9). *Assume that $\alpha_k \geq 0$ satisfies, for all $k \geq 1$,*

$$\boxed{\alpha_{k-1} \geq \alpha_k + \mathcal{D}\alpha_k^2} \tag{17}$$

where $\mathcal{D} > 0$. Then, for all $k \geq 1$,

$$\boxed{\alpha_k \leq \frac{\alpha_0}{1 + c\mathcal{D}\alpha_0 k} \leq \frac{1}{c\mathcal{D}k} = O\left(\frac{1}{k}\right)} \tag{18}$$

where

$$c := \frac{2}{1 + \sqrt{1 + 4\mathcal{D}\alpha_0}}.$$

Proof. Note that

$$\alpha_{k-1} \geq \alpha_k + \mathcal{D}\alpha_k^2 \qquad \text{if and only if} \qquad \mathcal{D}\alpha_{k-1} \geq \mathcal{D}\alpha_k + (\mathcal{D}\alpha_k)^2$$

and apply Lemma 1.9 to the sequence $\mathcal{D}\alpha_k$. □

Convergence rates for function values as a consequence of Corollary 1.10.

Proposition 1.11. *Consider the sequences evolved by the* PP algorithm (4) *and suppose $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Assume also that*

$$\boxed{\mathcal{D}_0 := \sup\left\{\|x - y\| \;\middle|\; \max\left\{f(x), f(y)\right\} \leq f(x_0)\right\} < +\infty.}$$

Then, for all $k \geq 1$,

$$\boxed{f(x_{k-1}) - f_{opt} \geq f(x_k) - f_{opt} + \frac{\lambda_k}{\mathcal{D}_0^2}\left(f(x_k) - f_{opt}\right)^2 \geq f(x_k) - f_{opt} + \frac{\underline{\lambda}}{\mathcal{D}_0^2}\left(f(x_k) - f_{opt}\right)^2.}$$

Proof. From Proposition 1.2(b),

$$f(x_{k-1}) \geq f(x_k) + \lambda_k \|v_k\|^2. \tag{19}$$

Let $\overline{x} \in \mathbb{R}^n$ be a solution of (1); in particular $f_{opt} = f(\overline{x})$. From (8) and the definition of $\partial f(x_k)$, we have

$$f(x_k) - f_{opt} \leq \langle v_k, x_k - \overline{x}\rangle \leq \|v_k\|\|x_k - \overline{x}\| \leq \|v_k\|\mathcal{D}_0. \tag{20}$$

In the latter inequality we used that $f(\overline{x}) \leq f(x_0)$, $f(x_k) \leq f(x_0)$ (see (19)) and the definition of $\mathcal{D}_0$. The desired result now follows by combining (19) and (20). □

Theorem 1.12 (Convergence rate for function values as a consequence of Corollary 1.10)**.** *Consider the sequences evolved by the* PP algorithm *and assume that* $\lambda_k \geq \underline{\lambda} > 0$ *for all* $k \geq 1$*, and* $0 < \mathcal{D}_0 < \infty$*. Then, for all* $k \geq 1$*,*

$$\boxed{f(x_k) - f_{opt} \leq \frac{f(x_0) - f_{opt}}{1 + c\,\frac{\underline{\lambda}}{\mathcal{D}_0^2}\big(f(x_0) - f_{opt}\big)\,k} \leq \frac{\mathcal{D}_0^2}{c\underline{\lambda}\,k} = O\left(\frac{1}{k}\right)} \tag{21}$$

where

$$c := \frac{2}{1 + \sqrt{1 + 4\frac{\underline{\lambda}}{\mathcal{D}_0^2}\big(f(x_0) - f_{opt}\big)}}.$$

Proof. The proof follows from Proposition 1.11 and Corollary 1.10 (with $\mathcal{D} = \underline{\lambda}/\mathcal{D}_0^2$). □

2 Generalizations of Corollary 1.9 – potentially useful for high-order algorithms.

Lemma 2.1. *Let* $\varphi\colon \mathbb{R}_+ \to \mathbb{R}$ *be a strictly increasing differentiable convex function and assume that* $a, b > 0$ *are such that* $\varphi(a) \leq 0 < \varphi(b)$*. Then,*

$$\boxed{a \leq b - \frac{\varphi(b)}{\varphi'(b)}.} \tag{22}$$

Proof. Using the inequality of the subdifferential we obtain $0 \geq \varphi(a) \geq \varphi(b) + \varphi'(b)(a - b)$, which gives the desired inequality (because $\varphi'(b) > 0$). □

Lemma 2.2 (Nesterov)**.** *Assume that* $\alpha_k \geq 0$ *satisfies, for all* $k \geq 1$*, and for some* $0 < \theta < 1$*,*

$$\boxed{\alpha_{k-1} \geq \alpha_k + \alpha_k^{1+\theta}.}$$

Then, for all $k \geq 1$*,*

$$\boxed{\alpha_k \leq \frac{\alpha_0}{(1 + c\,\alpha_0^\theta k)^{1/\theta}} \leq \frac{1}{(c\,k)^{1/\theta}} = O\left(\frac{1}{k^{1/\theta}}\right)} \tag{23}$$

where

$$c := \frac{\theta}{1 + \alpha_0^\theta}.$$

Proof. First note that if $\alpha_k = 0$, then the result is trivial. Assume now that $\alpha_k > 0$. Define $\beta_k = \alpha_k^\theta$. In particular, we have $\alpha_k = \beta_k^{1/\theta}$. Then, from the assumption,

$$\beta_k^{\frac{1+\theta}{\theta}} + \beta_k^{\frac{1}{\theta}} - \beta_{k-1}^{\frac{1}{\theta}} \leq 0.$$

Define, for $t \geq 0$,

$$\varphi(t) = t^{\frac{1+\theta}{\theta}} + t^{\frac{1}{\theta}} - \beta_{k-1}^{\frac{1}{\theta}}.$$

Note that $\varphi'(t) = \left(\frac{1+\theta}{\theta}\right) t^{\frac{1}{\theta}} + \frac{1}{\theta} t^{\frac{1-\theta}{\theta}}$. Clearly, φ is strictly increasing, convex and differentiable. Moreover, $\varphi(\beta_k) \leq 0 < \varphi(\beta_{k-1}) = \beta_{k-1}^{\frac{1+\theta}{\theta}}$. Hence, using Lemma 2.1, we obtain

$$\begin{aligned}
\beta_k &\leq \beta_{k-1} - \frac{\beta_{k-1}^{\frac{1+\theta}{\theta}}}{\left(\frac{1+\theta}{\theta}\right) \beta_{k-1}^{\frac{1}{\theta}} + \frac{1}{\theta} \beta_{k-1}^{\frac{1-\theta}{\theta}}} \\
&= \beta_{k-1} - \frac{\theta \beta_{k-1}^{\frac{1}{\theta}+1}}{(1+\theta)\, \beta_{k-1}^{\frac{1}{\theta}} + \beta_{k-1}^{\frac{1-\theta}{\theta}}} \\
&= \beta_{k-1} \left(1 - \frac{\theta \beta_{k-1}^{\frac{1}{\theta}}}{(1+\theta)\, \beta_{k-1}^{\frac{1}{\theta}} + \beta_{k-1}^{\frac{1-\theta}{\theta}}} \right) \\
&= \beta_{k-1} \left(\frac{\beta_{k-1}^{\frac{1}{\theta}} + \beta_{k-1}^{\frac{1-\theta}{\theta}}}{(1+\theta)\, \beta_{k-1}^{\frac{1}{\theta}} + \beta_{k-1}^{\frac{1-\theta}{\theta}}} \right) \\
&= \beta_{k-1} \left(\frac{1 + \beta_{k-1}^{-1}}{(1+\theta) + \beta_{k-1}^{-1}} \right) \\
&= \beta_{k-1} \left(\frac{\beta_{k-1} + 1}{\theta \beta_{k-1} + \beta_{k-1} + 1} \right).
\end{aligned}$$

So

$$\begin{aligned}
\frac{1}{\beta_k} &\geq \frac{1}{\beta_{k-1}} \left(\frac{\theta \beta_{k-1}}{\beta_{k-1} + 1} + 1 \right) \\
&= \frac{1}{\beta_{k-1}} + \frac{\theta}{\beta_{k-1} + 1}.
\end{aligned}$$

Using the above inequality and the fact that β_k is nonincreasing (i.e., that $\beta_0 \geq \beta_{k-1}$) we obtain

$$\begin{aligned}\frac{1}{\beta_k} \geq \frac{1}{\beta_{k-1}} + \frac{\theta}{\beta_0 + 1} &\geq \frac{1}{\beta_0} + \left(\frac{\theta}{\beta_0 + 1}\right) k \\ &= \frac{1}{\beta_0} + c\, k \\ &= \frac{1 + c\,\beta_0\, k}{\beta_0}.\end{aligned}$$

Hence,

$$\beta_k \leq \frac{\beta_0}{1 + c\beta_0 k},$$

which combined with the fact that $\beta_k = \alpha_k^\theta$, gives the desired result. □

Corollary 2.3 (Generalization of Lemma 2.2). *Assume that $\alpha_k \geq 0$ satisfies, for all $k \geq 1$, and for some $0 < \theta < 1$,*

$$\boxed{\alpha_{k-1} \geq \alpha_k + \mathcal{D}\alpha_k^{1+\theta}.}$$

Then, for all $k \geq 1$,

$$\boxed{\alpha_k \leq \frac{\alpha_0}{(1 + c\,\mathcal{D}\,\alpha_0^\theta k)^{1/\theta}} \leq \frac{1}{(c\,\mathcal{D}\,k)^{1/\theta}} = O\left(\frac{1}{k^{1/\theta}}\right)} \tag{24}$$

where

$$c := \frac{\theta}{1 + \mathcal{D}\,\alpha_0^\theta}.$$

Proof. Note that $\alpha_{k-1} \geq \alpha_k + \mathcal{D}\alpha_k^{1+\theta}$ if and only if $\mathcal{D}^{1/\theta}\alpha_{k-1} \geq \mathcal{D}^{1/\theta}\alpha_k + \left[\mathcal{D}^{1/\theta}\alpha_k\right]^{1+\theta}$. Hence, the proof follows by applying Lemma 2.2 to the sequence $\mathcal{D}^{1/\theta}\alpha_k$. □

Remark 2.4 (A remark on the PP algorithm)**.**

Proposition 2.5. *Let*

$$x_+ = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda}\|x - x_0\|^2 \right\}.$$

Then

(a) *If f is convex, then $f(x_0) \geq f(x_+) + \frac{1}{\lambda}\|x_+ - x_0\|^2$.*

(b) $f(x_0) \geq f(x_+) + \frac{1}{2\lambda}\|x_+ - x_0\|^2$.

Proof. (a) Note that, for all $x \in \mathbb{R}^n$, by the convexity of f,

$$f(x) + \frac{1}{2\lambda}\|x - x_0\|^2 \geq f(x_+) + \frac{1}{2\lambda}\|x_+ - x_0\|^2 + \frac{1}{2\lambda}\|x - x_+\|^2.$$

We now have just to take $x = x_0$.

(b) For all $x \in \mathbb{R}^n$, we have

$$f(x) + \frac{1}{2\lambda}\|x - x_0\|^2 \geq f(x_+) + \frac{1}{2\lambda}\|x_+ - x_0\|^2.$$

The desired inequality now follows by taking $x = x_0$. □

Note that (a) gives a better bound than (b), since (a) uses the convexity of f.

3 A duality view of the PP (HPE) method for convex optimization

Suppose we have $k \geq 1$ points in the graph of (the maximal monotone operator) ∂f:

$$\boxed{v_j \in \partial f(y_j), \qquad j = 1, \ldots, k.} \tag{25}$$

What could we do in this case?

Two options:

(i) Define *separating* hyperplanes ending up with *projective algorithms*.

(ii) Define methods via *duality* (through affine minorants).

In what follows we will work with option (ii) above. The starting point is to note that (25) yields, for all $x \in \mathbb{R}^n$,

$$\boxed{f(x) \geq \underbrace{f(y_j) + \langle v_j, x - y_j\rangle}_{=:\gamma_j(x)}, \qquad j = 1, \ldots, k.} \tag{26}$$

Now define the **aggregated functionals**:

$$\boxed{\Gamma_0 = 0, \qquad \Gamma_k = \textstyle\sum_{j=1}^k \lambda_j \gamma_j \qquad (k \geq 1).} \tag{27}$$

From the definition of Γ_k:

$$\boxed{\Gamma_{k+1} = \Gamma_k + \lambda_{k+1}\gamma_{k+1}, \quad k \geq 0.} \tag{28}$$

Note that $\nabla\gamma_j = v_j$ (for all $j \geq 1$). Also, define, for $k \geq 1$,

$$\boxed{x_k = \operatorname{argmin}_{x\in\mathbb{R}^n} \left\{\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2\right\},} \tag{29}$$

where $x_0 \in \mathbb{R}^n$, as well as (the dual variable)

$$\boxed{\beta_k = \inf_{x\in\mathbb{R}^n} \left\{\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2\right\}.} \tag{30}$$

Note that $\beta_0 = 0$.

The following result regarding x_k and β_k will be useful.

Lemma 3.1. *For all $k \geq 0$,*

$$\boxed{\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 = \beta_k + \frac{1}{2}\|x - x_k\|^2.}$$

Proof. Note first that if $k = 0$, then the result follows from the fact that $\Gamma_0 = 0$ and $\beta_0 = 0$. Assume now that $k > 0$. Note that $\varphi_k(x) := \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2$ is a quadratic function and $\nabla\varphi_k(x) = \nabla\Gamma_k(x) + x - x_0$ and (since Γ_k is affine) $\nabla^2\varphi_k(x) = I$. From (29) and (30), we find $\varphi_k(x_k) = \Gamma_k(x_k) + \frac{1}{2}\|x_k - x_0\|^2 = \beta_k$ and $\nabla\varphi_k(x_k) = 0$. Hence, using Taylor's theorem:

$$\begin{aligned}\varphi_k(x) &= \varphi_k(x_k) + \langle \underbrace{\nabla\varphi_k(x_k)}_{=0}, x - x_k\rangle + \frac{1}{2}\langle\nabla^2\varphi_k(x - x_k), x - x_k\rangle \\ &= \beta_k + \frac{1}{2}\|x - x_k\|^2,\end{aligned}$$

which combined with the definition of φ_k gives the desired result. □

Remark 3.2. *From* (25) *and* (27), *we get*

$$\Gamma_k(x) \leq \left(\sum_{j=1}^{k} \lambda_j\right) f(x). \tag{31}$$

Motivation. First note that from (31) we have a sort of weak duality:

$$\beta_k \leq \beta_k + \frac{1}{2}\|x - x_k\|^2 = \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \leq \left(\sum_{j=1}^{k} \lambda_j\right) f(x) + \frac{1}{2}\|x - x_0\|^2. \tag{32}$$

Hence,

We have that β_k is a lower-bound for the optimal value of f. In order to approximate the optimal value of f, it is enough to increase the value of β_k. But increase it to what proportion?

Let us try the following (strong duality):

$$\sum_{j=1}^{k} \lambda_j f(y_j) \leq \beta_k = \inf_{x\in\mathbb{R}^n} \left\{ \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\}$$
$$\leq \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2$$
$$\leq \left(\sum_{j=1}^{k} \lambda_j \right) f(x) + \frac{1}{2}\|x - x_0\|^2.$$

Summarizing, the strategy is as follows:

(i) To increase (the dual function) β_k.

(ii) Do (i) in a way that, at least,

$$\beta_k \geq \sum_{j=1}^{k} \lambda_j f(y_j).$$

Note that a sufficient condition for (ii) above is

$$\beta_{k+1} \geq \beta_k + \lambda_{k+1} f(y_{k+1}),$$

that is, it increases at least at the rate $\lambda_{k+1} f(y_{k+1})$.

That said, the next goal is to study the behaviour of (the dual function) β_k. Next lemma will help in this direction.

Lemma 3.3. *The following holds:*

(a) *For all* $k \geq 0$,

$$\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 = \beta_k + \lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2.$$

(b) *For all* $k \geq 0$,

$$x_{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ \lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2 \right\}.$$

(c) *For all* $k \geq 0$,

$$\beta_{k+1} = \beta_k + \inf_{x\in\mathbb{R}^n} \left\{ \lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2 \right\}.$$

Remark 3.4. *From item (a), we have (without aggregates!)*

$$\beta_{k+1} = \inf_{x\in\mathbb{R}^n} \left\{ \lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2 \right\} \qquad \text{for all } k \geq 0. \tag{33}$$

Moreover, the analogous of item (c) has not been observed for the accelerated method. I have to understand better this point here!

Proof. (a) Note that

$$\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \overset{(28)}{=} \Gamma_k(x) + \lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2$$

$$\overset{(a)}{=} \beta_k + \frac{1}{2}\|x - x_k\|^2 + \lambda_{k+1}\gamma_{k+1}(x).$$

(b) This follows from (a) and the definition of x_{k+1}; see (29).

(c) This follows from taking infimum in both sides of (a) and using the definition of β_{k+1}; see (30). □

Remark 3.5. *Lemma 3.3(c) motivates the study of the quantity, which gives the possible increasing behaviour of β_k,*

$$\boxed{\inf_{x\in\mathbb{R}^n}\left\{\lambda_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2\right\}.}$$

Note that, using the definition of $\gamma_{k+1}(\cdot)$ in (25), we have that the above infimum can be written as

$$\lambda_{k+1}f(y_{k+1}) + \inf_{x\in\mathbb{R}^n}\left\{\langle\lambda_{k+1}v_{k+1}, x - y_{k+1}\rangle + \frac{1}{2}\|x - x_k\|^2\right\}.$$

Lemma 3.6. *We have,*

$$\inf_{x\in\mathbb{R}^n}\left\{\langle v, x - y\rangle + \frac{1}{2}\|x - z\|^2\right\} = \frac{1}{2}\Big[\|y - z\|^2 - \|v + y - z\|^2\Big].$$

Proof. Note that

$$\inf_{x\in\mathbb{R}^n}\left\{\langle v, x - y\rangle + \frac{1}{2}\|x - z\|^2\right\} = \inf_{x\in\mathbb{R}^n}\left\{\langle v, x - z\rangle + \frac{1}{2}\|x - z\|^2\right\} + \langle v, z - y\rangle$$

$$= -\frac{\Delta}{4a} + \langle v, z - y\rangle$$

$$= -\frac{b^2}{4a} + \langle v, z - y\rangle$$

$$= -\frac{\|v\|^2}{4\frac{1}{2}} + \langle v, z - y\rangle$$

$$= -\frac{1}{2}\|v\|^2 + \langle v, z - y\rangle$$

$$= -\frac{1}{2}\Big[\|v\|^2 + 2\langle v, y - z\rangle + \|y - z\|^2\Big] + \frac{1}{2}\|y - z\|^2$$

$$= \frac{1}{2}\Big[\|y - z\|^2 - \|v + y - z\|^2\Big].$$

□

As a consequence of the above lemma and Lemma 3.3(d) we obtain:

Proposition 3.7. *The following holds:*

(a) *For all* $k \geq 0$,

$$\boxed{\beta_{k+1} = \beta_k + \lambda_{k+1} f(y_{k+1}) + \frac{1}{2}\Big[\|y_{k+1} - x_k\|^2 - \|\lambda_{k+1} v_{k+1} + y_{k+1} - x_k\|^2\Big].}$$

(b) *For all* $k \geq 1$,

$$\boxed{\beta_k = \textstyle\sum_{j=1}^{k} \lambda_j f(y_j) + \frac{1}{2}\sum_{j=1}^{k}\Big[\|y_j - x_{j-1}\|^2 - \|\lambda_j v_j + y_j - x_{j-1}\|^2\Big].}$$

(c) *For all* $k \geq 1$,

$$\boxed{\beta_k + \frac{1}{2}\|x - x_k\|^2 = \frac{1}{2}\|x - x_k\|^2 + \textstyle\sum_{j=1}^{k} \lambda_j f(y_j) + \frac{1}{2}\sum_{j=1}^{k}\Big[\|y_j - x_{j-1}\|^2 - \|\lambda_j v_j + y_j - x_{j-1}\|^2\Big].}$$

(d) *For all* $k \geq 1$,

$$\boxed{\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 = \frac{1}{2}\|x - x_k\|^2 + \textstyle\sum_{j=1}^{k} \lambda_j f(y_j) + \frac{1}{2}\sum_{j=1}^{k}\Big[\|y_j - x_{j-1}\|^2 - \|\lambda_j v_j + y_j - x_{j-1}\|^2\Big].}$$

Proof. (a) Follows from the fact that $\gamma_{k+1}(x) = f(y_{k+1}) + \langle v_{k+1}, x - y_{k+1}\rangle$, Lemma 3.3(c) and Lemma 3.6.

(b) By summing the equation in (a) for $j = 0, \ldots, k$ we obtain, for all $k \geq 1$,

$$\begin{aligned}\beta_k - \beta_0 = \sum_{j=0}^{k-1} [\beta_{j+1} - \beta_j] &= \sum_{j=0}^{k-1} \left(\lambda_{j+1} f(y_{j+1}) + \frac{1}{2}\Big[\|y_{j+1} - x_j\|^2 - \|\lambda_{j+1} v_{j+1} + y_{j+1} - x_j\|^2\Big]\right) \\ &= \sum_{j=1}^{k} \left(\lambda_j f(y_j) + \frac{1}{2}\Big[\|y_j - x_{j-1}\|^2 - \|\lambda_j v_j + y_j - x_{j-1}\|^2\Big]\right).\end{aligned}$$

The desired result now follows by using the fact that $\beta_0 = 0$.

(c) This follows directly by adding $\frac{1}{2}\|x - x_k\|^2$ in both sides of (b).

(d) This follows from (c) and Lemma 3.1. □

Proposition 3.8. *For all* $k \geq 1$,

$$\boxed{\|x - x_0\|^2 \geq \|x - x_k\|^2 + \textstyle\sum_{j=1}^{k} 2\lambda_j \left[f(y_j) - f(x)\right] + \sum_{j=1}^{k}\Big[\|y_j - x_{j-1}\|^2 - \|\lambda_j v_j + y_j - x_{j-1}\|^2\Big].}$$

Proof. This follows directly from Proposition 3.7(d) and (31). □

Remark 3.9. *It is possible to prove* Proposition 3.8 *without the aggreage functionals! Indeed, note that from* (33)*, we have*

$$\beta_k = \inf_{x\in\mathbb{R}^n}\left\{\gamma_k(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2\right\} \qquad \text{for all } k \geq 1.$$

Hence,

$$\begin{aligned}\gamma_k(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 &= \beta_k + \frac{1}{2\lambda_k}\|x - x_k\|^2 \\ &= f(y_k) + \inf_{x\in\mathbb{R}^n}\left\{\langle v_k, x - y_k\rangle + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2\right\} + \frac{1}{2\lambda_k}\|x - x_k\|^2.\end{aligned}$$

Since $\gamma_k \leq f$*, it follows that*

$$f(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 \geq f(y_k) + \frac{1}{2\lambda_k}\left[\|y_k - x_{k-1}\|^2 - \|\lambda_k v_k + y_k - x_{k-1}\|^2\right] + \frac{1}{2\lambda_k}\|x - x_k\|^2.$$

After multiplying both sides by $2\lambda_k > 0$ *and using a standard telescopic serie argument we obtian the same inequality as in* Proposition 3.8*, and in a much simpler way. So what exacly is the point of using or not the aggregate funcionals? Maybe the main motivation is acceleration!*

4 Accelerated methods

The results of this section are drawn from [7] and [6].

We will now show how the duality view of the PP algorithm developed in Section 3 can be used to devise accelereted versions of the PP algorithm.

New variables. Here, we will

generate two sequences of points $(x_k)_{k\geq 0}$ and $(y_k)_{k\geq 0}$, to be defined later.

The key ingredients:

1 A sequence of lower approximations by affine functionals $\gamma_k(\cdot)$ for function f: for all $k \geq 1$,

$$\boxed{\gamma_k(x) \leq f(x)} \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{34}$$

2 A sequence of scalars $a_k > 0$ for all $k \geq 1$.

3 A sequence of dual/aggregate functionals: $\boxed{\Gamma_0 \coloneqq 0}$ and, for all $k \geq 1$,

$$\boxed{\Gamma_k(x) \coloneqq \sum_{j=1}^{k} \frac{a_j}{A_k} \gamma_j(x) \leq f(x)} \quad \text{for all} \quad x \in \mathbb{R}^n \tag{35}$$

where, for all $k \geq 1$,

$$\boxed{A_k \coloneqq \sum_{j=1}^{k} a_j.} \tag{36}$$

Note that Γ_k as in (35) is an average!

4 A sequence $(\beta_k)_{k\geq 0}$ of dual approximations: For all $k \geq 0$,

$$\boxed{\beta_k \coloneqq \inf_{x\in\mathbb{R}^n} \left\{ A_k \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\} \leq A_k f(x) + \frac{1}{2}\|x - x_0\|^2} \quad \text{for all} \quad x \in \mathbb{R}^n \tag{37}$$

where $\boxed{A_0 \coloneqq 0}$. Note that $\boxed{\beta_0 = 0}$.

5 A sequence (y_k) in $\mathbb{R}^n$ such that $\boxed{A_k f(y_k) \leq \beta_k}$, i.e.,

$$\boxed{A_k f(y_k) \leq \inf_{x\in\mathbb{R}^n} \left\{ A_k \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\}.} \tag{38}$$

6 A lower bound $\boxed{A_k \geq O(k^p)}$ for all $k \geq 1$. For instance, for first-order methods, $p = 2$.

Note that (37) and (38) yield

$$\boxed{f(y_k) - f(x) \leq \underbrace{\frac{1}{2A_k}}_{\text{Rate!}} \|x - x_0\|^2} \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{39}$$

Remark 4.1. *The plan to achieve the goal in ingredient number 5, namely*

$$A_k f(y_k) \leq \beta_k \qquad \textit{for all } k \geq 0 \tag{40}$$

is to study the variation of these two sequences and prove that

$$\beta_{k+1} - \beta_k \geq A_{k+1} f(y_{k+1}) - A_k f(y_k) \quad \textit{for all} \quad k \geq 0.$$

Since these two function start at the same value, namely, $\beta_0 = A_0 f(y_0) = 0$, a simple telescopic argument is enough to prove (40).

Note that from the definition in (35) we obtain $A_k\Gamma_k = \sum_{j=1}^{k} a_j\gamma_j$ (for all $k \geq 1$) and so $A_{k+1}\Gamma_{k+1} = \sum_{j=1}^{k+1} a_j\gamma_j = \sum_{j=1}^{k} a_j\gamma_j + a_{k+1}\gamma_{k+1} = A_k\Gamma_k + a_{k+1}\gamma_{k+1}$. Using the facts that $\boxed{A_0 = 0}$ and $\boxed{\Gamma_0 = 0}$ we then obtain $A_{k+1}\Gamma_{k+1} = A_k\Gamma_k + a_{k+1}\gamma_{k+1}$ for all $k \geq 0$, i.e.,

$$\boxed{\Gamma_{k+1} = \frac{A_k}{A_{k+1}}\Gamma_k + \frac{a_{k+1}}{A_{k+1}}\gamma_{k+1} \quad \forall k \geq 0} \tag{41}$$

Note also that

$$\boxed{A_{k+1} = A_k + a_{k+1} \quad \forall k \geq 0.} \tag{42}$$

The primal variable x_k. Now define the sequence (x_k) as follows:

$$\boxed{x_k = \operatorname{argmin}_{x\in\mathbb{R}^n} \left\{ A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\} \qquad k \geq 0.} \tag{43}$$

The dual values β_k:

$$\boxed{\beta_k = \inf_{x\in\mathbb{R}^n} \left\{ A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 \right\} \qquad k \geq 0.} \tag{44}$$

Recall that $\boxed{\beta_0 = 0}$.

Variation of the dual values β_k.

Lemma 4.2. *The following holds for all $x \in \mathbb{R}^n$:*

(a) *For all $k \geq 0$,*

$$\boxed{A_k\Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2 = \beta_k + \frac{1}{2}\|x - x_k\|^2.}$$

(b) *For all $k \geq 0$,*

$$\boxed{A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 = \beta_k + a_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2.} \tag{45}$$

Generating $\gamma_k(\cdot)$ by computing points in the graph of the ε-subdifferential. Assume that we have

$$\boxed{v_j \in \partial_{\varepsilon_i} f(y_j), \qquad j = 1, \ldots, k+1.} \tag{46}$$

In particular, for $j = k+1$, we have

$$\boxed{v_{k+1} \in \partial_{\varepsilon_{k+1}} f(y_{k+1}).} \tag{47}$$

The computation of the pairs (y_j, v_j) will be specified later, through some proximal-point procedure!

The affine minorants. The inclusions in (46) yield, for all $j = 1, \ldots, k+1$,

$$\boxed{f(x) \geq \underbrace{f(y_j) + \langle v_j, x - y_j \rangle - \varepsilon_j}_{=:\gamma_j(x)} \qquad \text{for all } x \in \mathbb{R}^n.} \tag{48}$$

In particular, for $j = k+1$, we have

$$\boxed{f(x) \geq \gamma_{k+1}(x) \qquad \text{for all } x \in \mathbb{R}^n.} \tag{49}$$

An immediate consequence of (43), the definition of $\gamma_{k+1}(\cdot)$ as in (48), which gives in particular that $\boxed{\nabla \gamma_{k+1} = v_{k+1}}$, and (45) is as follows:

$$\boxed{x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ a_{k+1}\gamma_{k+1}(x) + \frac{1}{2}\|x - x_k\|^2 \right\}, \qquad k \geq 0}$$

which yields

$$\boxed{x_{k+1} = x_k - a_{k+1} v_{k+1}, \qquad k \geq 0.} \tag{50}$$

Next we will add the term $A_k f(y_k)$ in both sides of (45), since the sum $A_k = \sum_{j=1}^k a_j$ will give the improved complexity in function values.

Lemma 4.3. *For all $k \geq 0$, for all $x \in \mathbb{R}^n$,*

$$\boxed{A_k f(y_k) + A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \geq \beta_k + \underbrace{a_{k+1}\gamma_{k+1}(x) + A_k \gamma_{k+1}(y_k)}_{(**)} + \frac{1}{2}\|x - x_k\|^2.} \tag{51}$$

Proof. Adding $A_k f(y_k)$ in both sides of (45) we obtain

$$A_k f(y_k) + A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 = \beta_k + a_{k+1}\gamma_{k+1}(x) + A_k f(y_k) + \frac{1}{2}\|x - x_k\|^2.$$

To finish the proof, note that from (49) with $x = y^k$ we have $f(y_k) \geq \gamma_{k+1}(y_k)$ and so (after multiplying both sides of the latter inequality by A_k):

$$A_k f(y_k) \geq A_k \gamma_{k+1}(y_k).$$

□

The new variable $\widetilde{x}_k$; the middle-point in the segment.

The term $(**)$ in (51) motivates the definition of the following variables:

$$\widetilde{x}_k := \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x_k, \tag{52}$$

$$\widetilde{x} := \frac{A_k}{A_{k+1}} y_k + \frac{a_{k+1}}{A_{k+1}} x, \tag{53}$$

where $x \in \mathbb{R}^n$.

Note that the above definitions of $\widetilde{x}_k$ and $\widetilde{x}$ yield $\widetilde{x} - \widetilde{x}_k = \frac{a_{k+1}}{A_{k+1}}(x - x_k)$, i.e.,

$$x - x_k = \frac{A_{k+1}}{a_{k+1}}(\widetilde{x} - \widetilde{x}_k). \tag{54}$$

Note also that since $\gamma_{k+1}(\cdot)$ is affine and (53) is a convex combination (because $A_{k+1} = A_k + a_{k+1}$), we obtain $\gamma_{k+1}(\widetilde{x}) = \frac{A_k}{A_{k+1}}\gamma_{k+1}(y_k) + \frac{a_{k+1}}{A_{k+1}}\gamma_{k+1}(x)$, i.e.,

$$\boxed{A_{k+1}\gamma_{k+1}(\widetilde{x}) = \underbrace{A_k\gamma_{k+1}(y_k) + a_{k+1}\gamma_{k+1}(x)}_{\text{See } (**) \text{ in Lemma 4.3}}.} \tag{55}$$

Note now that direct substitution of (54) and (55) into (51) gives:

Lemma 4.4. *For all $k \geq 0$, for all $x \in \mathbb{R}^n$,*

$$\boxed{A_k f(y_k) + A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \geq \beta_k + A_{k+1}\gamma_{k+1}(\widetilde{x}) + \underbrace{\frac{A_{k+1}^2}{2a_{k+1}^2}}_{(***)} \|\widetilde{x} - \widetilde{x}_k\|^2.} \tag{56}$$

Proof. As we mentioned before, the proof follows from direct substitution of (54) and (55) into (51). □

We will now introduce a new variable:

$$\lambda_k > 0 \qquad \text{for all } \; k \geq 1.$$

The relationship between the scalars λ_k and a_k. Motivated by $(***)$ in (56) we have now to impose the following relationship between λ_k and a_k: $\frac{A_{k+1}}{a_{k+1}^2} = \frac{1}{\lambda_{k+1}}$, i.e.,

$$\boxed{a_{k+1}^2 = A_{k+1}\lambda_{k+1}.} \tag{57}$$

Explicit formula for computing a_{k+1} from λ_{k+1} and A_k:

$$\boxed{a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1}A_k}}{2}, \qquad k \geq 0.} \tag{58}$$

Indeed, from (57) and the fact that $A_{k+1} = A_k + a_{k+1}$ we obtain $a_{k+1}^2 = (A_k + a_{k+1})\lambda_{k+1}$, which can be written as a quadratic equation: $a_{k+1}^2 - \lambda_{k+1}a_{k+1} - \lambda_{k+1}A_k = 0$.

From (56) and (57) we obtain the following:

Lemma 4.5. *For all $k \geq 0$, for all $x \in \mathbb{R}^n$,*

$$\boxed{A_k f(y_k) + A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \geq \beta_k + A_{k+1}\left[\underbrace{\gamma_{k+1}(\widetilde{x}) + \frac{1}{2\lambda_{k+1}}\|\widetilde{x} - \widetilde{x}_k\|^2}_{(****)}\right].} \tag{59}$$

Proof. The proof is a direct consequence of (56) and (57). □

We now need the following lemma:

Lemma 4.6. *We have*

$$\inf_{x\in\mathbb{R}^n}\left\{\underbrace{\langle v, x - y\rangle - \varepsilon}_{\text{"linear" part of } \gamma(x)} + \frac{1}{2\lambda}\|x - z\|^2\right\} \geq \frac{1}{2\lambda}\left[\|y - z\|^2 - \left(\|\lambda v + y - z\|^2 + 2\lambda\varepsilon\right)\right].$$

As a consequence,

$$\|\lambda v + y - z\|^2 + 2\lambda\varepsilon \leq \sigma^2\|y - z\|^2 \quad \Rightarrow \quad \inf_{x\in\mathbb{R}^n}\left\{\underbrace{\langle v, x - y\rangle - \varepsilon}_{\text{linear part of } \gamma(x)} + \frac{1}{2\lambda}\|x - z\|^2\right\} \geq \frac{1 - \sigma^2}{2\lambda}\|y - z\|^2.$$

Proof. Note that

$$\begin{aligned}
\langle v, x-y\rangle - \varepsilon + \frac{1}{2\lambda}\|x-z\|^2 &= \frac{1}{\lambda}\left[\langle \lambda v, x-y\rangle + \frac{1}{2}\|x-z\|^2 - \lambda\varepsilon\right] \\
&= \frac{1}{\lambda}\left[\langle \lambda v, x-z\rangle + \frac{1}{2}\|x-z\|^2 - \lambda\varepsilon + \langle \lambda v, z-y\rangle\right] \\
&\geq \frac{1}{\lambda}\left[-\|\lambda v\|\|x-z\| + \frac{1}{2}\|x-z\|^2 - \lambda\varepsilon + \langle \lambda v, z-y\rangle\right] \\
&\geq \frac{1}{\lambda}\left[-\frac{\|\lambda v\|^2}{2} - \lambda\varepsilon - \langle \lambda v, y-z\rangle\right] \\
&= -\frac{1}{2\lambda}\left[\|\lambda v + y - z\|^2 - \|y-z\|^2 + 2\lambda\varepsilon\right] \\
&= \frac{1}{2\lambda}\left[\|y-z\|^2 - \left(\|\lambda v + y - z\|^2 + 2\lambda\varepsilon\right)\right].
\end{aligned}$$

□

Computing the variables (y_k, v_k) in the graph of $\partial_\varepsilon f$ as in (46).

First recall that (see (48)) $\boxed{\gamma_{k+1}(x) = f(y_{k+1}) + \langle v_{k+1}, x - y_{k+1}\rangle - \varepsilon_{k+1}}$. Motivated by $(***)$ and Lemma 4.6, we now know that we have to set:

$$\boxed{y_{k+1} \approx \operatorname{argmin}_{x\in\mathbb{R}^n}\left\{\lambda_{k+1} f(x) + \frac{1}{2}\|x - \widetilde{x}_k\|^2\right\}, \qquad k \geq 0} \tag{60}$$

in the following sense (HPE step):

$$\boxed{v_{k+1} \in \partial_{\varepsilon_{k+1}} f(y_{k+1}), \quad \|\lambda_{k+1} v_{k+1} + y_{k+1} - \widetilde{x}_k\|^2 + 2\lambda_{k+1}\varepsilon_{k+1} \leq \sigma^2 \|y_{k+1} - \widetilde{x}_k\|^2.} \tag{61}$$

From (61) and Lemma 4.6 we then obtain (motivation: see $(***)$ in (59))

$$\begin{aligned}
\inf_{x\in\mathbb{R}^n}\left\{\gamma_{k+1}(x) + \frac{1}{2\lambda_{k+1}}\|x-\widetilde{x}_k\|^2\right\} &= f(y_{k+1}) + \inf_{x\in\mathbb{R}^n}\left\{\langle v_{k+1}, x - y_{k+1}\rangle - \varepsilon_{k+1} + \frac{1}{2\lambda_{k+1}}\|x-\widetilde{x}_k\|^2\right\} \\
&\geq f(y_{k+1}) + \frac{1-\sigma^2}{2\lambda_{k+1}}\|y_{k+1} - \widetilde{x}^k\|^2,
\end{aligned}$$

which gives, in particular, that

$$\boxed{\gamma_{k+1}(\widetilde{x}) + \frac{1}{2\lambda_{k+1}}\|\widetilde{x} - \widetilde{x}_k\|^2 \geq f(y_{k+1}) + \frac{(1-\sigma^2)}{2\lambda_{k+1}}\|y_{k+1} - \widetilde{x}_k\|^2.} \tag{62}$$

Now from (59) and (62) we obtain the following:

Lemma 4.7. *For all $k \geq 0$, for all $x \in \mathbb{R}^n$,*

$$\boxed{A_k f(y_k) + A_{k+1}\Gamma_{k+1}(x) + \frac{1}{2}\|x - x_0\|^2 \geq \beta_k + A_{k+1} f(y_{k+1}) + \frac{(1-\sigma^2)A_{k+1}}{2\lambda_{k+1}}\|y_{k+1} - \widetilde{x}_k\|^2.} \tag{63}$$

Proof. The proof is a direct consequence of (59) and (62). □

FINALLY! The invariance!.

Lemma 4.8. *For all $k \geq 0$,*

$$\boxed{A_k f(y_k) + \beta_{k+1} \geq \beta_k + A_{k+1} f(y_{k+1}) + \frac{(1-\sigma^2)A_{k+1}}{2\lambda_{k+1}}\|y_{k+1} - \widetilde{x}_k\|^2.} \tag{64}$$

Proof. The proof follows by taking the infimum over $x \in \mathbb{R}^n$ in the left-hand side of (63) and then using the definition of β_{k+1} in (44). □

Recall that

$$\boxed{A_0 = 0.} \tag{65}$$

The above assumption will be important in the next lemma, essentially to ensure that $A_0 f(y_0) = 0$.

Lemma 4.9. *For all $k \geq 1$,*

$$\boxed{\beta_k \geq A_k f(y_k) + (1-\sigma^2)\sum_{j=1}^{k} \frac{A_j}{2\lambda_j}\|y_j - \widetilde{x}_{j-1}\|^2.} \tag{66}$$

Proof. From (64), for all $k \geq 0$,

$$\sum_{j=0}^{k}[\beta_{j+1} - \beta_j] \geq \sum_{j=0}^{k}[A_{j+1}f(y_{j+1}) - A_j f(y_j)] + (1-\sigma^2)\sum_{j=0}^{k}\left[\frac{A_{j+1}}{2\lambda_{j+1}}\|y_{j+1} - \widetilde{x}_j\|^2\right],$$

which gives, for all $k \geq 0$,

$$\beta_{k+1} - \underbrace{\beta_0}_{=0} \geq A_{k+1}f(y_{k+1}) - \underbrace{A_0 f(y_0)}_{=0;\,\text{See (65)}} + (1-\sigma^2)\sum_{j=0}^{k}\left[\frac{A_{j+1}}{2\lambda_{j+1}}\|y_{j+1} - \widetilde{x}_j\|^2\right],$$

which is clearly equivalent to (66). □

THE ALGORITHM! We have yet all the ingredients to build an algorithm. Namely, it will come from the conditions

$$\boxed{(65),\ (58),\ (52),\ (61),\ (42)\quad \text{and}\quad (50).}$$

Algorithm 1. Accelerated inexact PP algorithm (Monteiro-Svaiter)

(0) Choose $x_0, y_0 \in \mathbb{R}^n$ and $\sigma \in [0,1)$, let $A_0 = 0$ and set $k = 0$.

(1) Compute $\lambda_{k+1} > 0$ and $(y_{k+1}, v_{k+1}, \varepsilon_{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ such that

$$v_{k+1} \in \partial_{\varepsilon_{k+1}} f(y_{k+1}), \quad \|\lambda_{k+1} v_{k+1} + y_{k+1} - \widetilde{x}_k\|^2 + 2\lambda_{k+1}\varepsilon_{k+1} \le \sigma^2 \|y_{k+1} - \widetilde{x}_k\|^2$$

where

$$\widetilde{x}_k = \frac{A_k}{A_k + a_{k+1}} y_k + \frac{a_{k+1}}{A_k + a_{k+1}} x_k,$$

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1} A_k}}{2}.$$

(2) Let

$$A_{k+1} = A_k + a_{k+1},$$
$$x_{k+1} = x_k - a_{k+1} v_{k+1}.$$

(3) Set $k = k + 1$ and go to step 1.

We now proceed to compute convergence rates/iteration-complexity of Algorithm 1.

The iteration-complexity of Algorithm 1.

Lemma 4.10. *For all $k \ge 1$, for all $x \in \mathbb{R}^n$,*

$$\boxed{A_k f(y_k) + (1-\sigma^2)\sum_{j=1}^k \frac{A_j}{2\lambda_j}\|y_j - \widetilde{x}_{j-1}\|^2 + \frac{1}{2}\|x - x_k\|^2 \le A_k \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2.} \tag{67}$$

Proof. By adding $\frac{1}{2}\|x - x_k\|^2$ in both sides of (66) we obtain

$$A_k f(y_k) + \sum_{j=1}^k \frac{A_j}{2\lambda_j}\|y_j - \widetilde{x}_{j-1}\|^2 + \frac{1}{2}\|x - x_k\|^2 \le \beta_k + \frac{1}{2}\|x - x_k\|^2.$$

To finish the proof, note that from Lemma 4.2 we have

$$\beta_k + \frac{1}{2}\|x - x_k\|^2 = A_k \Gamma_k(x) + \frac{1}{2}\|x - x_0\|^2.$$

□

In order to proceed, we have now to prove the following property:

$$\boxed{A_k\Gamma_k \leq A_k f, \qquad k \geq 0} \tag{68}$$

which follows directly from the facts that $\Gamma_k \leq f$ for all $k \geq 1$ and $A_0 = 0$.

Theorem 4.11 (see [6]). *For all $k \geq 1$,*

$$\boxed{A_k\left[f(y_k) - f_{opt}\right] + (1-\sigma^2)\sum_{j=1}^{k} \frac{A_j}{2\lambda_j}\|y_j - \widetilde{x}_{j-1}\|^2 + \frac{1}{2}\|x_* - x_k\|^2 \leq \frac{d_0^2}{2}.} \tag{69}$$

As a consequence, for all $k \geq 1$,

$$\boxed{f(y_k) - f_{opt} \leq \frac{d_0^2}{2A_k}, \qquad \sum_{j=1}^{k} \frac{A_j}{\lambda_j}\|y_j - \widetilde{x}_{j-1}\|^2 \leq \frac{d_0^2}{1-\sigma^2} \quad \text{and} \quad \|x_* - x_k\| \leq d_0.} \tag{70}$$

Proof. Using (68), we obtain the following upper bound for the right-hand side of (67) with $x = x_*$:

$$A_k\Gamma_k(x_*) + \frac{1}{2}\|x_* - x_k\|^2 \leq A_k f_{opt} + \frac{d_0^2}{2},$$

which in turn combined with (67) with $x = x_*$ yields (69). To finish the proof, note that (70) follows readly from (69). □

Theorem 4.11 tell us, in particular, that for obtaining a convergence rate for

$$\boxed{f(y_k) - f_{opt}}$$

we have to

$$\boxed{\text{Study how the quantitie } A_k \text{ grows!}}$$

Recall that $\boxed{A_0 = 0}$ and, for $k \geq 0$,

$$\boxed{A_{k+1} = A_k + a_{k+1}}$$

where

$$\boxed{a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1}A_k}}{2}.}$$

Lemma 4.12 (see [6]). *The following holds:*

(a) *For all* $k \geq 0$, $A_{k+1} \geq \left(\sqrt{A_k} + \frac{\sqrt{\lambda_{k+1}}}{2}\right)^2$.

(b) *For all* $k \geq 0$, $\sqrt{A_{k+1}} - \sqrt{A_k} \geq \frac{\sqrt{\lambda_{k+1}}}{2}$.

(c) *For all* $k \geq 1$, $\boxed{A_k \geq \frac{1}{4}\left(\sum_{j=1}^{k} \sqrt{\lambda_j}\right)^2}$.

(d) *For all* $k \geq 1$, $\sum_{j=1}^{k} \|y^j - \widetilde{x}^{j-1}\|^2 \leq 4d_0^2$.

Proof. (a) Note first that

$$a_{k+1} = \frac{\lambda_{k+1} + \sqrt{\lambda_{k+1}^2 + 4\lambda_{k+1}A_k}}{2} \geq \frac{\lambda_{k+1}}{2} + \sqrt{\lambda_{k+1}A_k}.$$

Hence,

$$\begin{aligned} A_{k+1} = A_k + a_{k+1} &\geq A_k + \left(\frac{\lambda_{k+1}}{2} + \sqrt{\lambda_{k+1}A_k}\right) \\ &= \left[\left(\sqrt{A_k}\right)^2 + 2\sqrt{A_k}\frac{\sqrt{\lambda_{k+1}}}{2} + \left(\frac{\sqrt{\lambda_{k+1}}}{2}\right)^2\right] + \frac{\lambda_{k+1}}{4} \\ &= \left(\sqrt{A_k} + \frac{\sqrt{\lambda_{k+1}}}{2}\right)^2 + \frac{\lambda_{k+1}}{4} \\ &\geq \left(\sqrt{A_k} + \frac{\sqrt{\lambda_{k+1}}}{2}\right)^2, \end{aligned}$$

which gives (a).

(b) Note that (a) is clearly equivalent to $\sqrt{A_{k+1}} \geq \sqrt{A_k} + \frac{\sqrt{\lambda_{k+1}}}{2}$, which is of course equivalent to the statement in (b).

(c) Note that, from (b),

$$\sqrt{A_k} - \underbrace{\sqrt{A_0}}_{=0} = \sum_{j=1}^{k}\left(\sqrt{A_j} - \sqrt{A_{j-1}}\right) \geq \frac{1}{2}\sum_{j=1}^{k}\sqrt{\lambda_j},$$

which is clearly equivalent to (c).

(d) Note first that from (c) we obtain $A_k \geq \frac{\lambda_k}{4}$, i.e., $\boxed{\frac{A_k}{\lambda_k} \geq \frac{1}{4}}$. The desired result now follows from the latter inequality and the second inequality in (70). □

What is relevant here is not just the "format" of the aggregate functions $\Gamma_k(\cdot)$, but instead the possibility of making A_k large enough!

4.1 Bregman methods

Definition 4.13 (Bregman distance)**.**

$$\boxed{D_\varphi(x,y) = \varphi(x) - \varphi(y) - \langle \varphi'(y), x - y\rangle, \quad \forall x, y \in \mathbb{R}^n.} \tag{71}$$

Note that

$$\boxed{\frac{\partial}{\partial x} D_\varphi(x,y) = \varphi'(x) - \varphi'(y), \quad \forall x, y \in \mathbb{R}^n.} \tag{72}$$

Motivation:

$$\boxed{\|\underbrace{x^* - x}_{a}\|^2 - \|\underbrace{x^* - x^+}_{b}\|^2 = \|\underbrace{x^+ - x}_{a-b}\|^2 + \langle \underbrace{x^+ - x}_{a-b}, \underbrace{x^* - x^+}_{b}\rangle.} \tag{73}$$

Lemma 4.14. *The following holds:*

$$\boxed{D_\varphi(x^*, x) - D_\varphi(x^*, x^+) = D_\varphi(x^+, x) + \langle \varphi'(x^+) - \varphi'(x), x^* - x^+\rangle.} \tag{74}$$

Proof. Note that

$$\begin{aligned} D_\varphi(x^*, x) - D_\varphi(x^*, x^+) &= [\varphi(x^*) - \varphi(x) - \langle \varphi'(x), x^* - x\rangle] - [\varphi(x^*) - \varphi(x^+) - \langle \varphi'(x^+), x^* - x^+\rangle] \\ &= \varphi(x^+) - \varphi(x) - \langle \varphi'(x), x^* - x\rangle + \langle \varphi'(x^+), x^* - x^+\rangle \\ &= \varphi(x^+) - \varphi(x) - \langle \varphi'(x), x^+ - x\rangle + \langle \varphi'(x^+) - \varphi'(x), x^* - x^+\rangle \\ &= D_\varphi(x^+, x) + \langle \varphi'(x^+) - \varphi'(x), x^* - x^+\rangle. \end{aligned}$$

□

4.1.1 The Bregman PP algorithm

Consider the PP iteration: For $k \geq 0$,

$$\boxed{x^{k+1} = \operatorname{Arg\,min}_{x \in \mathbb{R}^n} \left\{\lambda_{k+1} f(x) + D_\varphi(x, x^k)\right\}, \quad \lambda_{k+1} > 0} \tag{75}$$

which is equivalent to, for all $k \geq 0$,

$$\boxed{v^{k+1} := -\frac{\partial}{\partial x} D_\varphi(x^{k+1}, x^k) = \frac{\varphi'(x^k) - \varphi'(x^{k+1})}{\lambda_{k+1}} \in \partial f(x^{k+1}).} \tag{76}$$

Concluding remarks and references. The regularization procedure described in (3) was first introduced by A. N. Tikhonov in the context of *ill-posed* operator equations and is now widely known as Tikhonov regularization (see, e.g., [12]). The function that maps z to the minimization problem (3)'s optimal value is referred to as the Moreau-Yosida regularization of f (see, e.g., [11]).

The proximal-point algorithm originates from the work of Martinet [4] and was further developed and popularized by Rockafellar's seminal contributions [10]. Much of the material presented here is drawn from [3], [5], and [2]. Specifically, Lemma 1.4 is from [3], while Lemma 1.5 and Proposition 1.7 are from [5]. Additionally, Lemma 1.9 is taken from [2] (see also [8]).

In modern research, the proximal-point algorithm serves as a foundation for designing and analyzing practical numerical schemes with theoretical performance guarantees. Notable examples include the proximal gradient method [8], the proximal-Newton method [2, 6] and, more recently, high-order tensor methods (see, e.g., [1, 9]), along with accelerated versions of theses methods (see, e.g., [6, 8]).

References

[1] M. M. Alves. Variants of the A-HPE and large-step A-HPE algorithms for strongly convex problems with applications to accelerated high-order tensor methods. *Optimization Methods and Software*, 37(6):2021–2051, 2022.

[2] M. M. Alves and B. F. Svaiter. A proximal-Newton method for unconstrained convex optimization in Hilbert spaces. *Optimization*, 67(1):67–82, 2018.

[3] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

[4] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.

[5] R. D. C. Monteiro and B. F. Svaiter. Convergence rate of inexact proximal point methods with relative error criteria for convex optimization. Optimization online technical report (url: https://optimization-online.org/?p=11242), 2010.

[6] R. D. C. Monteiro and B. F. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *SIAM J. Optim.*, 23(2):1092–1125, 2013.

[7] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152, 2005.

[8] Y. Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, Cham, 2018. Second edition of [MR2142598].

[9] Y. Nesterov. Implementable tensor methods in unconstrained convex optimization. *Math. Program.*, 186(1-2, Ser. A):157–183, 2021.

[10] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[11] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[12] A. N. Tikhonov and V. Ya. Arsenin. *Methods of Ill-Posed Problems Solving*. Science, Moscow, 1979.