Dynamical adjustment of the prox-parameter in bundle methods

Pablo A. Rey* Claudia Sagastizábal†

Abstract

Proximal bundle methods are well known for their efficiency in nondifferentiable optimization. Their interpretation as approximate proximal methods yields a certain reversal poor-man formula for the regularization parameter in the proximal term. A new updating rule for this prox-parameter is introduced, based on the same scheme, but making use of all the information available at each iteration. Numerical results assessing the validity of the approach are reported.

1 Introduction. Motivation

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a convex lower semicontinuous function. The problem of finding a minimizer of f can be solved by various nonsmooth optimization methods such as: subgradients ([39]), cutting-planes ([5], [18]), analytic centers ([10]), and bundle methods.

Since the pioneering works [40], [23], bundle methods have been intensively studied in an effort towards obtaining robust variants with good convergence rates. The fundamental idea is to stabilize the somewhat erratic behaviour of cutting-planes algorithms. Given the bundle $\{y^i, f(y^i), g^i \in \partial f(y^i)\}_{i \le j}$ collecting information of previous iterations, the new cutting-plane iterate, y^{j+1}, solves the j-th subproblem

$$\min_{y \in \mathbb{R}^N} \check{f}(y) \quad \text{with } \check{f}(y) := \max_{i \le j} f(y^i) + \left\langle g^i, y - y^i \right\rangle . \tag{1}$$

A bundle stabilization consists in modifying (1) in order to prevent oscillations, for instance as follows:

*PUC, Depto. Engenharia Elétrica. Rio de Janeiro, RJ, CEP 22453-900, Brazil. Research supported by CNPq under Bolsa de Doutorado processo No.142835/97-8.

†IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. On leave from INRIA-Rocquencourt, BP 105, 78153 Le Chesnay, France. Research partially supported by CNPq under Grant No. 301800/96-0 and FAPERJ under Grant No. 150-205/98.

– Trust-region stabilization: as in [38], where the minimization in (1) is done on a certain ball of radius Δ.
– Level stabilization: as in [25], a "dual" point of view, where a certain radius is minimized subject to decrease $\check{f}$ below a level ℓ.
– Penalization: as in [19], where a a quadratic term with penalty parameter μ is added to the criterion in (1). This algorithm is also called *proximal bundle method.*

We refer to [12, Ch. XV.2.1] for a detailed description of the resulting algorithms.

It is known, see for instance [4, Theorem 7.7], that all the stabilizations above produce conceptually equivalent subproblems. In practice, however, a different management of the parameters (Δ, μ, ℓ) results in different sequences and, thus, in different numerical behaviour. Indeed, in relation with speeds of convergence, a proper tuning of parameters appears as a crucial issue.

The question of how to choose the prox-parameter is important not only for bundle methods, but also in other fields of optimization. Such is the case of Tikhonov's regularization methods, penalty methods coupled with the proximal algorithm ([21], [22]) and various smoothing techniques for optimization and variational problems ([31], [15]). For example, in [17] an appropriate choice of μ ensures convexity of subproblems when the objective is nonconvex. Similarly, in [16] proximal barrier algorithms are controlled via a dynamical adjustment of the proximal term that preserves self-concordance of the barrier.

This paper focuses on a particular instance of proximal bundle methods, called *Reversal Quasi-Newton Bundle algorithm* (RQB) in [30]. For this type of methods, fast convergence is related to an appropriate decrease of the prox-parameter μ. The new updating rule that we present here aims at producing a certain smallest value of μ. While the updating formula in RQB is based on only one (sub)gradient, our rule makes use of all the available subgradient information. To choose the best value for μ we solve (analytically) a minimization problem over conic sections, i.e., general quadratic equations.

Our paper is organized as follows. Next section recalls the basics of proximal bundle methods, in particular RQB in its "poor-man" variant. In § 2.4 therein, we give our proposal for updating μ, expressed as an optimization problem with quadratic constraints. We give a modified RQB algorithm (**mRQB**) and assess our approach by reporting encouraging numerical results in § 3. Finally, we include at the end of the paper two appendices with auxiliary results needed in our development. Appendix A contains some elements of Analytical Geometry related to parameterizations of conic sections. In Appendix B we discuss the analytical resolution of the optimization problem associated to μ.

2 Proximal Bundle methods

In this section we explain in detail the penalized jth-subproblem solved by proximal bundle methods as well as the formula proposed in [30] to update the penalization parameter μ.

2.1 Some generalities

Let $\bar{x}$ denote a minimizer of f. It is known that cutting-planes methods are particularly suitable when $\partial f(\bar{x})$ is a polyhedron. For general f, however, subproblem (1) may become ill-conditioned when j grows (see [34] for an example). As a result, the method stalls and sometimes fails, due to numerical errors.

Bundle methods appear as a two steps stabilization of cutting-planes methods. First, they generate a sampling point y^{j+1} by solving a stabilized subproblem as in (2) below. Second, they *select* some iterates satisfying certain descent condition. These selected points, or *stability centers*[1], form a subsequence $\{x^n\} \subset \{y^j\}$ that aims at guaranteeing a sufficient decrease in f. Specifically, given the last generated stability center x^n and $\mu_n > 0$, y^{j+1} is the solution of

$$\min_{y \in \mathbb{R}^N} \check{f}(y) + \frac{1}{2}\mu_n \|y - x^n\|^2 . \tag{2}$$

If for some $m_1 \in]0, 1[$

$$f(y^{j+1}) \le f(x^n) - m_1 \left(f(x^n) - \check{f}(y^{j+1}) - \frac{1}{2}\mu_n \|y^{j+1} - x^n\|^2 \right) , \tag{3}$$

then y^{j+1} provides a sufficient decrease. A new stability center is thus defined, $x^{n+1} = y^{j+1}$, μ_n is updated, and both j and n are increased.

Otherwise, the model $\check{f}$ is considered not accurate enough. Then a *null step* is declared: there is no new stability center and only j is increased.

Note that in both cases a new affine function of the form $f(y^{j+1}) + \langle g^{j+1}, y - y^{j+1} \rangle$ is incorporated to the max-function in (1). The new bundle values $f(y^{j+1})$ and $g^{j+1} \in \partial f(y^{j+1})$ are obtained by calling an oracle or *black box*. Also, note that our notation doesn't make explicit j, the number of affine functions defining the maximum (1). This apparent lack of precision comes from the fact that only descent indices n will be of interest in our analysis.

As usual in nonlinear programming, the method can be globalized by merging the bundling technique above with a line-search. Since the line-search function is not differentiable, this may be a tricky issue. A possibility is to do a curve-search, like subalgorithm (CS) in [30], introducing a trial step-size $t > 0$ in the quadratic term in (2). We omit entering into more technicalities here. We refer to [38], [30] for more details.

We now give an interpretation of the scheme above in terms of the proximal algorithm, [33], [37].

[1] also called descent- or serious-steps in the bundle jargon.

2.2 The role of the Moreau-Yosida regularization

Recall that for a convex function f the Moreau-Yosida regularization $F(x)$ at a given point $x \in \mathbb{R}^N$ is the optimal value of

$$\min_{y \in \mathbb{R}^N} f(y) + \frac{1}{2}\mu\|y - x\|^2 \,, \tag{4}$$

where μ is a positive parameter. The corresponding unique minimizer is the proximal point of f at x, denoted by $p(x)$. Observe that the optimality condition of (4) yields the relation

$$p(x) = x - \frac{1}{\mu}G(p(x)) \,, \text{where } G(p(x)) \in \partial f(p(x)) \,. \tag{5}$$

Moreover,

$$G(p(x)) = \nabla F(x) = \mu(x - p(x)) \tag{6}$$

is a Lipschitz continuous function, [33],[37].

Remark 2.1 *It is also well known that minimizing f is equivalent to minimize F. In this respect, replacing x by x^n and $p(x)$ by x^{n+1} in (5) gives an iterative procedure for minimizing F. This procedure is a gradient method, preconditioned by the diagonal matrix $1/\mu\, I$. Its rate of convergence can be accelerated by means of a preconditioner as in quasi-Newton methods. We come back to this important point in § 2.3.* □

As observed in [1], [8], [6], a loop on j-indices without changing n in (2) (i.e., a sequence of null steps between two stability centers) can be interpreted as an implementable mechanism to compute the proximal point operator of f at x^n. Indeed, when there is sufficient decrease, $p(x^n)$ is approximated by x^{n+1} which, from the optimality condition of (2), can be expressed as follows:

$$(p(x^n) \approx) \quad x^{n+1} = x^n - \frac{1}{\mu_n}\check{G}(x^{n+1}) \,, \tag{7}$$

where

$$\check{G}(x^{n+1}) := \nabla \check{F}(x^n) = \mu_n(x^n - x^{n+1}) \in \partial \check{f}(x^{n+1}). \tag{8}$$

Thus, the resolution of subproblem (2) gives for free $\check{G}(x^{n+1})$, an approximated subgradient of f at x^{n+1}.

2.3 Reversal forms in bundle methods

Keeping in mind Remark 2.1, we see that (7) is one step for minimizing F using a gradient method. Ideally, the preconditioner factor $1/\mu_n\, I$ should approximate some second-order object of F at an optimum $\bar{x}$. For a thorough study on second-order developments of convex functions we refer to [28], [29], [26], [32] . We explain now how to use a quasi-Newton formula to update μ_n, called *poor man's* in [30]. Namely, given $(v,u) \in \mathbb{R}^N \times \mathbb{R}^N$, define:

$$\mu_{n+1} = \frac{\|v\|^2}{\langle v, u\rangle}. \tag{9}$$

The pair (v,u) is formed by some difference of gradients and points. A classical choice is

$$(v,u) = \left(\nabla F(x^{n+1}) - \nabla F(x^n), x^{n+1} - x^n\right).$$

However, we see from (6) that to compute $\nabla F(x^{n+1})$ an additional calculation approximating $p(x^{n+1})$ is required.

Instead, we use the *reversal form*, introduced in [30], that we describe next. After x^{n+1} is defined, the black box gives a subgradient $g^{n+1} \in \partial f(x^{n+1})$. The idea is to find a point z^{n+1} such that $x^{n+1} = p(z^{n+1})$, using (5)-(6). More precisely, writing (5) with $p(x), x, G(p(x))$ replaced, respectively, by $x^{n+1}, z^{n+1}, g^{n+1}$ yields

$$x^{n+1} = z^{n+1} - \frac{1}{\mu_n} g^{n+1} \quad \Longrightarrow \quad z^{n+1} = x^{n+1} + \frac{1}{\mu_n} g^{n+1}. \tag{10}$$

Observe that a different subgradient in $\partial f(x^{n+1})$ would give a different z^{n+1}. A similar calculation at x^n using x^n, z^n, g^n gives

$$\begin{aligned}
(v,u) &= \Big(\nabla F(z^{n+1}) - \nabla F(z^n) \quad , \quad z^{n+1} - z^n \Big) \\
&= \Big(\mu_n(z^{n+1} - p(z^{n+1})) - \mu_n(z^n - p(z^n)) \quad , \quad z^{n+1} - z^n \Big) \\
&= \Big(\mu_n(z^{n+1} - x^{n+1}) - \mu_n(z^n - x^n) \quad , \quad z^{n+1} - z^n \Big).
\end{aligned}$$

Together with (10), we obtain

$$(v,u) = \left(g^{n+1} - g^n, x^{n+1} - x^n + \frac{1}{\mu_n}(g^{n+1} - g^n)\right). \tag{11}$$

For completeness we reproduce here the main steps of the algorithm proposed in [30, p. 405]. We omit detailing the curve-search subalgorithm (CS) therein, because it is not relevant for our purposes. However, in the numerical experiments reported in §§ 3.2 we do use the full RQB-CS scheme from [30].

Reversal quasi-Newton Bundle Algorithm (RQB)

STEP 0. The initial x^1 and μ_1 are given, as well as a tolerance $0 < m_1 < 1$. Set $y^1 = x^1$. Define the initial cutting-plane model $\check{f}_1$ according to (1) and set $n = j = 1$.

STEP 1. Solve (2) and call y^{j+1} its solution. Compute $f(y^{j+1})$ and $g^{j+1} \in \partial f(y^{j+1})$. Decide if y^{j+1} is

– either a null-step, when (3) does not hold;

– or a descent-step, when (3) holds.

STEP 2. In case of null-step, go to Step 3. Otherwise update $x^{n+1} = y^{j+1}$ and update μ_{n+1} as in (9) and (11). Replace n by $n + 1$.

STEP 3. Update the cutting-plane model $\check{f}$ of (1). Replace j by $j + 1$; loop to Step 1. $\square$

Concerning convergence results, in Theorems 7 and 8 of [27] it is shown that the sequence $\{x^n\}$ generated by RQB is either finite with last element optimal, or infinite and minimizing (i.e., $f(x^n) \to \inf f$ as $n \to \infty$). When f is strictly convex a superlinear convergence result holds. Specifically, letting $\bar{\mu}$ be the limit point of decreasing sequence $\{\mu_n\}$, Theorem 4.4. in [30] establishes the following:

– If $\bar{\mu} = 0$ then RQB is superlinearly convergent.

– If $\bar{\mu} > 0$ then the sequence $\{f(x^n)\}$ is 2-step superlinearly convergent.

2.4 Towards richer poor-man's formulæ

The reversal formula (11) is used in Step 2 of RQB to update μ as in (9). The rate of convergence results mentioned above as well as empirical observation (see [30, p. 399]) indicate that fast convergence of the algorithm is related to an appropriate decrease of the prox-parameter, μ. The analysis we present in the sequel aims at producing a certain smallest value for μ in (9).

We have seen that the value of v depends on the subgradients considered. Also, once the difference of subgradients v has been chosen, the difference of points u is determined in a straightforward manner from (11). Since any difference between two subgradients (at x^{n+1} and x^n) can be used to define v, we choose among all the differences available, the one providing the least value for μ.

Recall that there are two subgradients available at each point x^n:

– First, there is g^n, the subgradient given by the oracle.

– Second, from (6), there is $G(x^n)$ the gradient of the Moreau-Yosida regularization at x^{n-1}. Although the exact value in (6) is not known, (8) gives the following approximation:

$$G^n := \check{G}(x^n) \approx \nabla F(x^{n-1}) .$$

Likewise, at x^{n+1} we have g^{n+1} and $G^{n+1} := \check{G}(x^{n+1})$.

By convexity, the whole segment joining g^{n+1} and G^{n+1}

$$\left[\alpha_1 g^{n+1} + (1-\alpha_1)G^{n+1}\right], \quad \alpha_1 \in [0,1]$$

is contained in (an approximation of) $\partial f(x^{n+1})$ and the same holds at x^n for the segment joining g^n and G^n. We parameterize v as a function of the corresponding convex coefficients:

$$\begin{array}{rlll} v(\alpha) & :=[\alpha_1 g^{n+1} + (1-\alpha_1)G^{n+1}] - [\alpha_2 g^n + (1-\alpha_2)G^n] & & \\ & = (G^{n+1} - G^n) & +\alpha_1(g^{n+1} - G^{n+1}) & +\alpha_2(G^n - g^n) \\ & =: v^0 & +\alpha_1 v^1 & +\alpha_2 v^2\,, \end{array} \tag{12}$$

where $\alpha = (\alpha_1, \alpha_2) \in [0,1]^2$. Accordingly, we define

$$u(\alpha) := x^{n+1} - x^n + \frac{1}{\mu_n}\, v(\alpha)\,. \tag{13}$$

Corresponding to this parameterization, the smallest available value for μ_{n+1} is the optimal value of

$$\left\{ \begin{array}{l} \inf\limits_{\alpha\in[0,1]^2} \dfrac{\|v(\alpha)\|^2}{\langle u(\alpha), v(\alpha)\rangle} \\ \langle u(\alpha), v(\alpha)\rangle > 0\,, \end{array} \right. \tag{14}$$

where the objective is simply (9) and the positivity constraint is used to ensure $\mu > 0$. The feasible set in (14) is nonempty, since it contains $\alpha = (1,1)$. This is because $v(1,1) = g^{n+1} - g^n$ and, in this case, the positivity of $\langle u(1,1), v(1,1)\rangle$ is ensured by using a Wolfe-type condition in the curve-search algorithm CS (see [30, p.399]). In Figure 1 it can be seen that this feasible set is nonconvex, furthermore, it may even be non connected.

Using the equality $G^{n+1} = \mu_n(x^n - x^{n+1})$ in (13) yields the relation $u(\alpha) = (v(\alpha) - G^{n+1})/\mu_n$. With this expression, (14) is equivalent, up to a factor $1/\mu_n$, to the problem

$$\left\{ \begin{array}{l} \inf\limits_{\alpha\in[0,1]^2} \varphi(\alpha) := \dfrac{\|v(\alpha)\|^2}{\langle v(\alpha), v(\alpha) - G^{n+1}\rangle} \\ \langle u(\alpha), v(\alpha)\rangle > 0\,. \end{array} \right. \tag{15}$$

Note that it is possible for the infimum of this problem to be zero. Indeed, there can be a feasible sequence $\{\alpha^n\}_n$ such that, letting $\alpha^* = \lim_{n\to\infty} \alpha^n$, it holds both $\lim_{n\to\infty} \varphi(\alpha^n) = 0$ and $\langle v(\alpha^*), u(\alpha^*)\rangle = 0$.

Nevertheless, we are interested in the value α giving the smallest value for φ in (15), but we want *feasible* α's. Moreover, from the implementation point of view, it would be convenient to bound this value away from zero. In order to achieve these goals, we impose the scalar product $\langle u, v\rangle$ to be "positive enough". That is to say, we force $\langle u, v\rangle$ in (15) to be greater than a fixed small parameter $\varepsilon > 0$:

$$\begin{cases} \min\limits_{\alpha\in[0,1]^2} \varphi(\alpha) = \dfrac{\|v(\alpha)\|^2}{\langle v(\alpha), v(\alpha) - G^{n+1}\rangle} \\ \langle u(\alpha), v(\alpha)\rangle \geq \varepsilon\,. \end{cases} \tag{16}$$

Let $\mathcal{F}^\varepsilon$ denote the feasible set in (16)

$$\mathcal{F}^\varepsilon := \left\{ \alpha \in [0,1]^2 : \langle u(\alpha), v(\alpha)\rangle \geq \varepsilon \right\}. \tag{17}$$

Because of the special form of $\mathcal{F}^\varepsilon$ (see Figure 1), to find an explicit solution of (16) we split this set in two parts, namely, its boundary and interior, as described in §§ B.1 and B.2 in the Appendix.

Observe that, from (13) and (12), $\langle u(\alpha), v(\alpha)\rangle = \varepsilon$ is a quadratic equation on α. More precisely, it describes a conic section, as defined in Appendix A below.

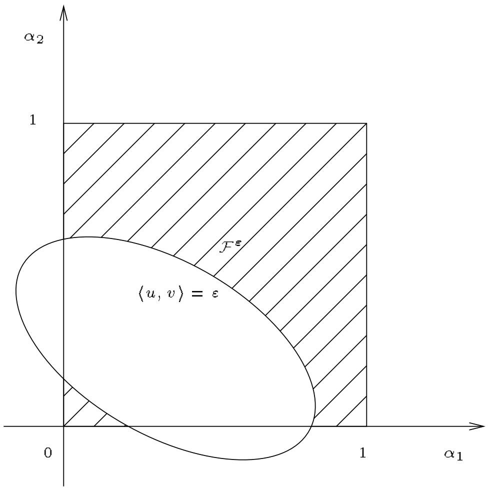

Figure 1: Feasible set of (16).

In general, v^1 and v^2 in (12) are linearly independent. Otherwise, if there is $\lambda \in \mathbb{R}$ such that $v^2 = \lambda v^1$, then (12) becomes $v(\alpha) = v^0 + (\alpha_1 + \alpha_2\lambda)v^1$. Thus, letting $t := \alpha_1 + \alpha_2\lambda$, both (12) and (13) are reparameterized as functions of $t \in \mathbb{R}$:

$$v(t) := v^0 + tv^1 \quad \text{and} \quad u(t) := x^{n+1} - x^n + \frac{1}{\mu_n}\, v(t)\,. \tag{18}$$

In this case, the following straightforward result holds.

Proposition 2.2 *If $v^2 = \lambda v^1$ for some scalar λ, then* (16) *can be reformulated as the following one-dimensional problem:*

$$\left\{ \begin{array}{l} \min\limits_{t \in I} \dfrac{\|v(t)\|^2}{\langle v(t), v(t) - G^{n+1}\rangle} \\ \langle u(t), v(t)\rangle \geq \varepsilon \end{array} \right. \tag{19}$$

where $v(t)$ and $u(t)$ are defined in (18) *and*

$$I := \left\{ \begin{array}{ll} [0, 1+\lambda] & \textit{if } \lambda \geq 0 \\ [\lambda, 1] & \textit{if } \lambda < 0\,. \end{array} \right.$$

□

The curved boundary $\langle u(\alpha), v(\alpha)\rangle = \varepsilon$ in (16) plays a key role in our development. From (12) and (13) we see that this is a quadratic equation in α. This motivates our review on general quadratic equations, or "conics", see Appendix A at the end of the paper. Appendix B contains necessary optimality conditions for (14) derived from the conic sections theory. These conditions define sets of stationary points containing the desired solution.

3 Numerical considerations

In this section we describe the actual modifications introduced in RQB using the analysis developed in Appendix B. We also report some experimental results.

3.1 Updating μ

We already mentioned that the update of the prox-parameter μ_n is done by taking the optimal value in (14). Up to a scalar factor, this comes to solve (15). In fact, we solve the implementable form (16) by examining its boundary and interior solutions. Among the various candidates or stationary points, we choose the one giving the best value of φ in (16) to calculate μ_{n+1}.

Boundary candidates are obtained using Proposition B.2, that we rewrite here for completeness. The proof can be found in Appendix B.1.

Proposition B.2 *If α is a solution of* (16) *that lies on $\mathcal{F}_b$, the boundary of the feasible set* (17), *the following holds:*

Either α lies on one side $S \in \{\{0\}\times[0,1], \quad \{1\}\times[0,1], \quad [0,1]\times\{0\}, \quad [0,1]\times\{1\}\}$. *Then α has one component equal to 0 or 1, while the other component solves*

$$a_2 h_0 t^2 + 2a_2 h_1 t + a_1 h_1 - a_0 h_0 = 0\,, \tag{29}$$

where the coefficients a_0, a_1, a_2 *and* h_0, h_1 *depend on the side* S *considered.*

Or α *lies on the curved part of the boundary, and it can be computed by solving a system of quadratic equations with a single unknown.*

Otherwise, α *is an extreme point in* $\{0,1\} \times \{0,1\}$. □

Interior candidates are found via Proposition B.3, proved in Appendix B.2:

Proposition B.3 *Let* A *be the* $N \times 3$ *matrix* $\left[\begin{array}{ccc} v^0 & v^1 & v^2 \end{array}\right]$. *For* $i = 1, 2$ *define the* 3×3 *symmetric matrices*

$$B^i := A^\top [\left\langle G^{n+1}, v^i \right\rangle I - (v^i {G^{n+1}}^\top + G^{n+1} {v^i}^\top)] A \,,$$

where I *is the identity matrix of order* N. *Then the solutions* $\alpha = (\alpha_1, \alpha_2)$ *of* (16) *that lie in the interior of the feasible set* (17) *are solutions of the quadratic system*

$$\left\{ \begin{array}{rcl} \langle B^1 \nu, \nu \rangle & = & 0 \\ \langle B^2 \nu, \nu \rangle & = & 0 \,, \end{array} \right. \tag{31}$$

where $\nu := (1, \alpha_1, \alpha_2)^\top$. □

3.2 The algorithm

To implement the dynamical adjustment explained in § 2.4, we just modify one step in the Reversal quasi-Newton Bundle Algorithm (RQB) from [30], that we reproduced at the end of § 2.3. Specifically, given a parameter $\varepsilon > 0$ at *Step 0*, we replace *Step 2* by

Step 2'

In case of null-step, go to *Step 3*. Otherwise, update $x^{n+1} = y^{j+1}$ and execute the following block

Step 2'.a **(test of linear independence)**

If v^1 and v^2 in (12) are linearly independent, go to *Step 2'.b*. Otherwise, solve (19).

Step 2'.b **(extreme points)**

Let $\mathcal{A} := \{0,1\} \times \{0,1\}$.

Step 2'.c **(boundary points)**

Apply Proposition B.2 as follows:

– solve (29) for each side S.

– Find the stationary points lying on the curved border.

– Add all the candidates to $\mathcal{A}$.

Step 2'.d **(interior points)**

Apply Proposition B.3 as follows:

– compute B^i in (31) and check for degeneracy.

If one of the conics is degenerate call it B^2 in (31).

Otherwise, find a real root of the cubic equation $\det(B^1 + \lambda B^2) = 0$.

If such a root does not exist go to *Step 2'.e*.

Otherwise, replace $\langle B^2\nu, \nu\rangle = 0$ in (31) by $\langle (B^1 + \lambda B^2)\nu, \nu\rangle = 0$.

– Factorize the equation of the (second) degenerate conic.

If the lines obtained have complex coefficients, go to *Step 2'.e*.

Otherwise, find all the points $\nu = (1, \alpha_1, \alpha_2)^\top$ in the intersection of these lines with the other conic section. Add such α to the set $\mathcal{A}$.

Step 2'.e **(update of μ_n)**

Compute $\varphi(\alpha)$ in (16) for all $\alpha \in [0,1]^2 \cap \mathcal{A}$.

Let φ^* be the least computed value. Define $\mu_{n+1} = \varphi^* \mu_n$.

Replace n by $n+1$. □

3.3 Numerical experiments

We assess the good performance of our approach versus RQB by testing both updating rules on a set of nondifferentiable problems. We call our updating rule **m**RQB, by **modified** RQB. It was implemented in Fortran as a subroutine of the package N1CV2, available upon request from `Claude.Lemarechal@inrialpes.fr`.

All runs were performed on a Pentium II 400MHz with 32Mb RAM. The parameter ε in (16) was set to 10^{-8} for all the runs. We mention in passing that we obtained almost identical results when varying $\varepsilon \in \{10^{-4}, 10^{-6}, 10^{-8}\}$.

Test examples include both academic and power-system problems. The group of academic problems is formed by

– `BADGUY`, dimension 30 and 100, a worst-case function, see [25];

– `MAXQUAD` and `TR48`, described in [24];

– `PCB442`, a dual of a 1-tree relaxation of a traveling salesman problem, see [11], [36];

– `TRUSS15.2`, a truss optimization problem with 15556 bars, see [20], [14]; and

– `NETxxxx`, 3 dual network problems, see [3], [9].

We use three models for each network problem: `NETa` denotes the aggregated variant while `NETb` and `NETc` denote different disaggregated variants taking advantage of the underlying separable structure. We refer to [2] for a description of disaggregated bundle methods.

We report the number of calls to the black box needed to achieve a precision of $10^{-3,-4,-6}$ in the objective function (the stopping test was deactivated). Tables 1 and 3 show the results for our academic and power-system examples, respectively. In a column titled "gain" we put + for those cases where our updating rule was better.

In relation with Table 1, some comments are in order. First, the resilient nature of `BADGUY` is confirmed once again. Small size problems, such as `MAXQUAD` or `TR48` do not seem to be improved by our modified updating rule. The beneficial effects of **m**`RQB` become

	10^{-3}			10^{-4}			10^{-6}		
	RQB	**m**RQB	gain	RQB	**m**RQB	gain	RQB	**m**RQB	gain
BADGUY30	95	95		126	125	+	187	185	+
BADGUY100	319	319		418	418		519	519	
MAXQUAD	49	93		52	93		75	93	
TR48	134	156		205	201	+	205	212	
PCB442	79	96		271	196	+	517	376	+
TRUSS15.2	13310	7557	+	13992	7568	+	14179	7572	+
NETa22L	661	654	+	691	682	+	756	834	
NETb22L	244	244		246	244	+	248	248	
NETc22L	248	248		252	252		254	252	+
NETa22H	!!	909	+	!!	1037	+	!!	1081	+
NETb22H	236	249		237	250		238	252	
NETc22H	238	246		239	248		241	252	
NETc148	163	191		237	233	+	561	334	+

Table 1: Academic examples. $\#f/g$ evaluations

more evident for higher accuracies and bigger problems (such as PCB442, NETc148). This observation will be confirmed by our power-system results of Table 3. It is also interesting to observe in our NETxxxx problems how different cutting-planes models (i.e., different aggregated/disaggregated bundle methods) determine the degree of success of both RQB and **m**RQB.

test #	thermic plants	time steps	nodes
1	10	24	129
2	10	48	881
3	10	72	2161
4	30	24	129
5	30	48	881
6	30	72	2161

Table 2: Power-system problems.

Our power system problems belong to the battery of stochastic unit-commitment problems described in [2, § 5]. There are 6 different problems having the configurations in Table 2.

The number of nodes in Table 2 corresponds to the dimension of the nondifferentiable

	10^{-3}			10^{-4}			10^{-6}		
	RQB	**mRQB**	gain	RQB	**mRQB**	gain	RQB	**mRQB**	gain
TEST1	41	49		114	124		173	184	
TEST1**p**	39	39		127	126	+	158	151	+
TEST2	33	33		83	83		546	546	
TEST2**p**	37	37		80	89		129	138	
TEST3	110	107	+	!!	426	+	!!	721	+
TEST3**p**	45	48		110	120		212	196	+
TEST4	51	50		156	163		316	247	+
TEST4**p**	23	23		207	164	+	340	298	+
TEST5	54	54		137	105	+	555	435	+
TEST5**p**	57	57		129	129		194	190	+
TEST6	157	148	+	373	329	+	!!	472	+
TEST6**p**	68	68		158	148	+	284	!!	

Table 3: Unit-commitment examples. $\#f/g$ evaluations

dual problem, that was solved by using a disaggregated bundle method. We ran each test problem without and with the ad-hoc preconditioning introduced in [2]. To distinguish both kind of runs, we write a **p** for the latter.

The runs in Table 3 confirm our previous comments: the higher the accuracy, the better **mRQB** is. We also observe that the efficiency of the preconditioning from [2] is not spoiled by **mRQB**. With respect to TEST6**p**, although we report a failure for our updating rule, if we activate the stopping test, both RQB and **mRQB** stop after 257 and 271 calls to the oracle, respectively. Such is not the case for RQB failures in TEST3,6.

We finish by saying that, even though our analysis to find the solutions of (16) is rather involved, the actual implementation *Step 2'* requires negligible extra computational times. An interesting feature of **mRQB**is its robust behaviour with respect to changes in ε. Altogether, we recommend **mRQB** for large-scale problems where high precision is desired.

Acknowledgments. We are indebted to the referees for a careful reading of this paper.

A APPENDIX: A view on Analytic Geometry

We introduce some notation and definitions.

– Given $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, let

$$\nu := \begin{bmatrix} 1 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \tag{20}$$

be a vector in $\mathbb{R}^3$.

– Given a 3×3 symmetric matrix B, with entries b_{ij}, the set

$$\left\{ \alpha \in \mathbb{R}^2 : \langle B\nu, \nu \rangle = 0\,, \text{ with } \nu \text{ as in (20)} \right\}, \tag{21}$$

is called a *conic section* [7, p.208]. Note that the equation in (21) is quadratic, with two unknowns. Indeed, any arbitrary quadratic equation with two unknowns can be written as in (21).

– The determinant of B in (21) is called the *discriminant* of the conic section.

– A *degenerate* conic, is a conic whose discriminant is null. In this case, the curve is composed by two lines and the equations of these two lines can be obtained by factorizing $\langle B\nu, \nu \rangle = 0$. Examples are given by the set $\{\alpha \in \mathbb{R}^2 : \alpha_1 \alpha_2 = 0\}$, composed by the two coordinate axes; as well as the conic obtained by taking

$$B = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

in (21), which gives the equation $(\alpha_1 - \alpha_2)^2 = 0$.

– A non degenerate conic section is a parabola when $b_{11} b_{22} - b_{12}^2 = 0$. In Remark B.1 we will see that parabolas can be omitted in our analysis.

– When the conic is an hyperbola, an ellipse or a circle, we can define its center (i.e., its symmetry center) and axes (a pair of orthogonal conjugated diameters).

Lemma A.1 *If the conic $\langle B\nu, \nu \rangle$ is an hyperbola or an ellipse then,*
(a) its center (α_1^0, α_2^0) is given by the unique *solution of the equations*

$$\begin{aligned} b_{11}\alpha_1 + b_{12}\alpha_2 + b_{13} &= 0 \\ b_{12}\alpha_1 + b_{22}\alpha_2 + b_{23} &= 0\,, \end{aligned}$$

(b) the slopes m of the (orthogonal) axes are the solutions of the quadratic equation

$$b_{12} m^2 + (b_{11} - b_{22}) m - b_{12} = 0\,. \tag{22}$$

Proof. See [35, pp.150 and 156].

□

A.1 Parameterizing a conic section

The parameterization of a conic section is involved and needs a detailed description. The simplest case to parameterize a conic is when B is diagonal. Consider a non singular diagonal matrix

$$B = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix}, \tag{23}$$

i.e., $b_{ii} \neq 0$ for all i. When all three elements b_{ii} have the same sign, the conic is imaginary (it has no real point), a case of no interest for our analysis. We suppose thus that there exists an element with different sign.

Lemma A.2 *Let B be a non singular diagonal matrix as in (23). Suppose further that two of its elements have one sign while the third one has the opposite sign. Define $\kappa_i = \frac{1}{\sqrt{|b_{ii}|}}$, for $i = 1, 2, 3$. Then the following holds:*

(a) If $-\operatorname{sign}(b_{11}) = \operatorname{sign}(b_{22}) = \operatorname{sign}(b_{33})$, the conic is an ellipse or a circle parameterized by

$$\begin{aligned} \alpha_1'(t) &= \frac{(1-t^2)\kappa_2}{(1+t^2)\kappa_1} \\ \alpha_2'(t) &= \frac{2t\kappa_3}{(1+t^2)\kappa_1}, \qquad \textit{for all } t \in \mathbb{R}\,. \end{aligned}$$

(b) If $\operatorname{sign}(b_{11}) = -\operatorname{sign}(b_{22}) = \operatorname{sign}(b_{33})$, the conic is an hyperbola parameterized by

$$\begin{aligned} \alpha_1'(t) &= \frac{(1+t^2)\kappa_2}{(1-t^2)\kappa_1} \\ \alpha_2'(t) &= \frac{2t\kappa_3}{(1-t^2)\kappa_1}, \qquad \textit{for all } t \notin [-1,1]\,. \end{aligned}$$

When $\operatorname{sign}(b_{11}) = \operatorname{sign}(b_{22}) = -\operatorname{sign}(b_{33})$, the parameterization is similar, replacing κ_2 by κ_3.

Proof. Without loss of generality, suppose that $b_{11} > 0$.
(a) We have $b_{22} < 0$, $b_{33} < 0$ and

$$\kappa_1 = \frac{1}{\sqrt{b_{11}}},\ \kappa_2 = \frac{1}{\sqrt{-b_{22}}},\ \kappa_3 = \frac{1}{\sqrt{-b_{33}}}.$$

Plugging the expressions above in the equation $\langle B\nu, \nu\rangle = 0$ we obtain

$$\frac{\alpha_1^2}{\kappa_2^2} + \frac{\alpha_2^2}{\kappa_3^2} = \frac{1}{\kappa_1^2} \iff \frac{\alpha_1^2}{\left(\frac{\kappa_2}{\kappa_1}\right)^2} + \frac{\alpha_2^2}{\left(\frac{\kappa_3}{\kappa_1}\right)^2} = 1\,.$$

Now recall the well-known parameterization,

$$\begin{array}{lcll} \alpha_1'(\theta) & := & \frac{\kappa_2}{\kappa_1}\cos\theta & \\ \alpha_2'(\theta) & := & \frac{\kappa_3}{\kappa_1}\sin\theta, & \qquad \text{for } \theta \in [0, 2\pi], \end{array}$$

and let $t := \tan\frac{\theta}{2}$. The result follows.
(b) We have $b_{22} < 0$, $b_{33} > 0$, and

$$\kappa_1 = \frac{1}{\sqrt{b_{11}}},\ \kappa_2 = \frac{1}{\sqrt{-b_{22}}},\ \kappa_3 = \frac{1}{\sqrt{b_{33}}}.$$

Plugging the expressions above in the equation $\langle B\nu, \nu\rangle = 0$ we obtain

$$\frac{\alpha_1^2}{\left(\frac{\kappa_2}{\kappa_1}\right)^2} - \frac{\alpha_2^2}{\left(\frac{\kappa_3}{\kappa_1}\right)^2} = 1 \Longleftrightarrow \alpha_2^2 = \left(\frac{\kappa_3}{\kappa_1}\right)^2 \left(\frac{\alpha_1^2}{\left(\frac{\kappa_2}{\kappa_1}\right)^2} - 1\right). \tag{24}$$

Note that points α on the hyperbola (24) are defined only when $|\alpha_1| \geq \frac{\kappa_2}{\kappa_1}$. As a result, there exists an angle θ such that $\cos\theta = (\frac{\kappa_2}{\kappa_1})/\alpha_1$. Now, substituting by $\alpha_1 := \frac{\kappa_2}{\kappa_1}\sec\theta$ in (24) we obtain

$$(\alpha_2)^2 = \left(\frac{\kappa_3}{\kappa_1}\right)^2 \left(\sec^2\theta - 1\right) = \left(\frac{\kappa_3}{\kappa_1}\right)^2 \tan^2\theta\,.$$

As in the first part, setting $t := \tan\frac{\theta}{2}$, the result follows.

□

Remark A.3 *In the case of the ellipse, the parameterization above leaves out the point* $\left(-\frac{\kappa_2}{\kappa_1}, 0\right)$*, corresponding to the limit value* $t = \infty$*. Thus, we consider this point separately in our analysis.* □

If B is not diagonal, then it can be diagonalized first by using some of the methods given in [13, pp.167-179] or by using a change of variable as we describe next. Geometrically, if B is diagonal, the center of the conic coincides with the origin of the coordinate system, while the coordinate axes are the axes of the conic. Hence, when B is not diagonal, it can by diagonalized by changing the coordinate system so that its center is the center of the conic and the same for the axes. For example, associate to B the matrix

$$Q := \begin{bmatrix} 1 & -m & \alpha_1^0 \\ m & 1 & \alpha_2^0 \\ 0 & 0 & 1 \end{bmatrix},$$

where (α_1^0, α_2^0) is the center of the conic and m is one of the solutions of (22). Then the matrix $Q^\top BQ$ is diagonal and represents the conic in the new coordinate system $(\tilde{\alpha}_1, \tilde{\alpha}_2)$:

$$\begin{aligned}
\tilde{\alpha}_1 &:= \frac{-m\alpha_2^0 + m\alpha_2 - \alpha_1^0 + \alpha_1}{1+m^2} \\
\tilde{\alpha}_2 &:= -\frac{\alpha_2^0 - m\alpha_1^0 + m\alpha_1 - \alpha_2}{1+m^2}.
\end{aligned}$$

After parameterizing the conic section in this new coordinate system we can go back to the original coordinate system by simply changing variables back to the original ones.

We now address the problem of finding analytical solutions of (16). Basically, we use the parameterizations above to obtain various necessary optimality conditions. These conditions define sets of stationary points containing the desired solution.

B APPENDIX: Finding stationary points

We have seen in § 2.4 that we can choose μ_{n+1} by solving (16). Here we discuss how to solve this problem. Keeping in mind Proposition 2.2, we assume in the sequel that v^1 and v^2 are linearly independent.

We already mentioned that in our analysis we split the feasible set $\mathcal{F}^\varepsilon$ into two parts: one intersecting the boundary and the other comprising its interior.

B.1 Boundary solutions

The boundary of the feasible set $\mathcal{F}^\varepsilon$ for the problem (16) can be decomposed in the sides of the square $[0,1]^2$, namely, the segments

$$\{0\} \times [0,1], \quad \{1\} \times [0,1], \quad [0,1] \times \{0\}, \quad [0,1] \times \{1\} \tag{25}$$

and the set

$$\left\{\alpha \in [0,1]^2 : \langle u(\alpha), v(\alpha)\rangle = \varepsilon\right\}, \tag{26}$$

that is a set of arcs of a conic section.

B.1.1 Sides of the square

Along each side S from (25), the problem (16) is one-dimensional, via the parameterization

$$[0,1] \ni t \longmapsto v = w^0 + tw^1, \tag{27}$$

where w^0 and w^1 depend on S. For instance, in the side $S = \{0\} \times [0,1]$, there is just one free component for $\alpha = (0,t)$. Then

$$v(\alpha) = v(t) = v^0 + tv^2,$$

which means $(w^0, w^1) := (v^0, v^2)$. For the remaining sides the corresponding pairs (w^0, w^1) are

$$(v^0 + v^1, v^2), \quad (v^0, v^1), \quad (v^0 + v^2, v^1)\,.$$

Therefore, instead of (16) we solve

$$\min_{t\in[0,1]} \frac{a_2t^2 + a_1t + a_0}{a_2t^2 + (a_1 - h_0)t + (a_0 - h_1)}\,, \tag{28}$$

where we have defined

$$a_0 := \|w^0\|^2\,,\ a_1 := 2\left\langle w^0, w^1\right\rangle\,,\ a_2 := \|w^1\|^2$$

and

$$h_i := \left\langle w^i, G^{n+1}\right\rangle \text{ for } i = 0, 1\,.$$

B.1.2 Curved boundary

The constraint set $\{\alpha : \langle u(\alpha), v(\alpha)\rangle = \varepsilon\}$ in (16) is a conic section. We now show how a suitable parameterization this set yields a unidimensional problem, equivalent to (16).

Given the $N \times 3$ matrix $A := \begin{bmatrix} v^0 & v^1 & v^2 \end{bmatrix}$, the Gram matrix $A^\top A$ has entries $\left\langle v^i, v^j\right\rangle$. With this notation, to express the constraint $\langle u(\alpha), v(\alpha)\rangle = \varepsilon$ as in (21), it is enough to take

$$B \;:=\; A^\top A - \begin{bmatrix} h_0 - \varepsilon & h_1/2 & h_2/2 \\ h_1/2 & 0 & 0 \\ h_2/2 & 0 & 0 \end{bmatrix},$$

where we have defined

$$h_i := \left\langle v^i, G^{n+1}\right\rangle \text{ for } i = 0, 1, 2\,.$$

Remark B.1 *Without loss of generality we can assume that B is non singular and such that $b_{11}b_{22} - b_{12}^2 \neq 0$ (this condition excludes parabolas).* □

The next proposition characterizes boundary solutions.

Proposition B.2 *If α is a solution of* (16) *that lies on $\mathcal{F}_b$, the boundary of the feasible set* (17)*, the following holds:*

Either α lies on one side S from (25)*. In this case, α has one component equal to 0 or 1, while the other component solves*

$$a_2h_0t^2 + 2a_2h_1t + a_1h_1 - a_0h_0 = 0\,, \tag{29}$$

where the coefficients a_0, a_1, a_2 and h_0, h_1 depend on the side S considered.

Or α lies on the curved part of the boundary, and it can be computed by solving a system of quadratic equations with a single unknown.

Otherwise, α is an extreme point in the set $\{0, 1\} \times \{0, 1\}$.

Proof. Recall that a minimizer of a one dimensional problem is either a stationary point or an extreme of (the intervals that compose) the feasible set. Candidates to solve (16), or stationary points, can be obtained by examining first, the parameterized sides of the square $[0,1]\times[0,1]$; second, the parameterized arcs of the conic $\{\alpha : \langle u(\alpha), v(\alpha)\rangle = \varepsilon\}$; and, finally, the extreme points of the square.

We start with the sides, enumerated in (25). Note that different choices of S correspond to different sides of the square. The corresponding pair (w^0, w^1) gives the parameterizations (27) of this side. For each side, differentiate the objective in (28) to obtain the stationary points lying on S.

Second, we analyze the curved boundary. Recalling Appendix A, let B be the (diagonal) matrix associated to the conic section $\{\alpha : \langle u(\alpha), v(\alpha)\rangle = \varepsilon\}$. Then Lemma A.2 gives an analytic expression for $\alpha = \alpha(t)$, $t \in \mathbb{R}$, depending on the sign of the elements b_{ii}. Since in this region the objective in (16) is the (affine) functional $\varphi(\alpha) = 1 + \dfrac{\langle G^{n+1}, v(\alpha)\rangle}{\varepsilon}$, the reparameterized objective, say $\psi(t) := \varphi(\alpha(t))$, is just a quotient of quadratic functions of t. Then stationary points of (16) can be obtained by differentiating $\psi(t)$ and zeroing the resulting quadratic numerator. $\square$

Now we consider the interior solutions of (14).

B.2 Interior solutions

Keeping the same notation, we can rewrite φ in (16) as

$$\varphi = \frac{\langle A\nu, A\nu\rangle}{\langle A\nu, A\nu\rangle - \langle A\nu, G^{n+1}\rangle}\,. \tag{30}$$

Proposition B.3 *For $i = 1,2$ let B^i be a 3×3 symmetric matrix defined by*

$$B^i := A^\top[\left\langle G^{n+1}, v^i\right\rangle I - (v^i {G^{n+1}}^\top + G^{n+1}{v^i}^\top)]A\,,$$

where I is the identity matrix of order N. Then the solutions of (16) *that lie in the interior of the feasible set are solutions of the quadratic system*

$$\begin{cases} \langle B^1\nu, \nu\rangle &= 0 \\ \langle B^2\nu, \nu\rangle &= 0\,. \end{cases} \tag{31}$$

Proof. To compute $\nabla_\alpha\varphi = 0$ we first differentiate φ from (30) with respect to ν. Call ϕ_1 and ϕ_2 the first and second term in the denominator in (30), respectively. Using the particular form of the quotient therein, we see that the stationary points of φ are solutions of

$$\begin{aligned} 0 &= \text{numerator of } \nabla_\nu\left(\frac{\phi_1}{\phi_1-\phi_2}\right) \\ &= \phi_1\nabla_\nu\phi_2 - \phi_2\nabla_\nu\phi_1 \\ &= \langle A\nu, A\nu\rangle A^\top G^{n+1} - 2\left\langle A\nu, G^{n+1}\right\rangle A^\top A\nu\,. \quad (*) \end{aligned}$$

Recalling (20), to obtain the derivatives with respect to α we only consider the two last components in (*):

$$0 = \langle A\nu, A\nu\rangle \left\langle v^i, G^{n+1}\right\rangle - 2\left\langle A\nu, G^{n+1}\right\rangle \left\langle v^i, A\nu\right\rangle \quad \text{for } i = 1, 2\,.$$

Now perform some simple algebraic manipulations to write the following equalities

$$\begin{aligned} 0 &= \left\langle A^\top A\nu, \nu\right\rangle \left\langle v^i, G^{n+1}\right\rangle - 2\left\langle A\nu, G^{n+1}\right\rangle \left\langle A^\top v^i, \nu\right\rangle \\ &= \left\langle A^\top(\left\langle v^i, G^{n+1}\right\rangle I)A\nu, \nu\right\rangle - 2\left\langle A^\top v^i {G^{n+1}}^\top A\nu, \nu\right\rangle \\ &= \left\langle A^\top(\left\langle v^i, G^{n+1}\right\rangle I - 2v^i {G^{n+1}}^\top)A\nu, \nu\right\rangle\,. \end{aligned}$$

The proof is finished, by taking the symmetric part of the matrix $2v^i{G^{n+1}}^\top$. □

Note that when (31) does not have solutions there are no interior candidates.

Also, note that each equation in the quadratic system (31) describes a conic section. Their intersection gives the interior candidates. The solution set in Proposition B.3 is not changed if we rewrite (31) so that one of the conics therein is degenerate (recall that a conic is said to be *degenerate* if its discriminant is zero). In this case, (21) can be expressed as the product of two lines. Thus, its intersection with the other conic in (31) requires the resolution of two quadratic equations. This is the procedure used in *Step 2'.d* of the modified `RQB` algorithm **mRQB** given in §§ 3.2.

References

[1] A. Auslender. Numerical methods for nondifferentiable convex optimization. *Mathematical Programming Study*, 30:102–126, 1987.

[2] L. Bacaud, C. Lemaréchal, A. Renaud, and C.A. Sagastizábal. Bundle methods in stochastic optimal power management: a disaggregated approach using preconditioners. *Computational Optimization and Applications*, 2000. Accepted for publication.

[3] D.P. Bertsekas and E.M. Gafni. Two-metric projection methods for constrained optimization. *SIAM Journal on Control and Optimization*, 22:936–964, 1984.

[4] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Optimisation Numérique: aspects théoriques et pratiques*. Springer Verlag, 1997.

[5] E. Cheney and A. Goldstein. Newton's method for convex programming and tchebycheff approximations. *Numerische Mathematik*, 1:253–268, 1959.

[6] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Mathematical Programming*, 62(2):261–275, 1993.

[7] L.P. Einsenhart. *Coordinate Geometry.* Dover Publications, 1960.

[8] M. Fukushima. A descent algorithm for nonsmooth convex optimization. *Mathematical Programming*, 30:163–175, 1984.

[9] J.-L. Goffin, J. Gondzio, R. Sarkissian, and J.-P. Vial. Solving nonlinear multicommodity flow problems by the analytic center cutting plane method. *Mathematical Programming*, pages 131–154, 1997.

[10] J.L. Goffin, A. Haurie, and J.Ph. Vial. Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38(2):284–302, 1992.

[11] M. Held and R. Karp. The travelling salesman problem and minimum spanning trees: Part II. *Mathematical Programming*, 1:6–25, 1971.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).

[13] C.M. Hoffmann. *Geometric and Solid Modeling: An Introduction.* Morgan Kaufmann, 1989.

[14] F. Jarre, M. Kočvara, and J. Zowe. Optimal truss design by interior-point methods. *SIAM Journal on Optimization*, 8:1084–1107, 1998.

[15] A. Kaplan and R. Tichatschke. Prox-regularization and solution of ill-posed elliptic variational inequalities. *Applied Mathematics*, 42(2):111–145, 1997.

[16] A. Kaplan and R. Tichatschke. Proximal methods in view of interior-point strategies. *Journal of Optimization Theory and Applications*, 98(2):399–429, 1998.

[17] A. Kaplan and R. Tichatschke. Proximal point methods and nonconvex optimization. *Journal of Global Optimization*, 13(4):389–406, 1998.

[18] J. E. Kelley. The cutting plane method for solving convex programs. *J. Soc. Indust. Appl. Math.*, 8:703–712, 1960.

[19] K.C. Kiwiel. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming*, 46:105–122, 1990.

[20] M. Kočvara and J. Zowe. How mathematics can help in design of mechanical structures. In D.F. Griffiths and G.A. Watson, editors, *Numerical Analysis 1995*, pages 76–93. Longman, Harlow, 1996.

[21] B. Lemaire. The proximal algorithm. In *New methods in optimization and their industrial uses*, number 87 in Internat. Schriftenreihe Numer. Math., pages 73–87. Birkhauser, Basel-Boston, MA, 1989.

[22] B. Lemaire. On the convergence of some iterative methods for convex minimization. In *Recent developments in optimization*, number 429 in Lecture Notes in Econom. and Math. Systems, pages 252–268. Springer-Verlag, Berlin, 1995.

[23] C. Lemaréchal. An extension of Davidon methods to nondifferentiable problems. *Mathematical Programming Study*, 3:95–109, 1975.

[24] C. Lemaréchal and R. Mifflin. A set of nonsmooth optmization test problems. In C. Lemaréchal and R. Mifflin, editors, *Nonsmooth Optimization*, pages 151–165. Pergamon Press, Oxford, 1978.

[25] C. Lemaréchal, A. Nemirovskii, and Yu. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–148, 1995.

[26] C. Lemaréchal, F. Oustry, and C. Sagastizábal. The $\mathcal{U}$-Lagrangian of a convex function. *Trans. Amer. Math. Soc.*, 352(2):711–729, 2000.

[27] C. Lemaréchal and C. Sagastizábal. An approach to variable metric bundle methods. In J. Henry and J-P. Yvon, editors, *Systems Modelling and Optimization*, number 197 in Lecture Notes in Control and Information Sciences, pages 144–162. Springer-Verlag. Also as RR INRIA #2128, 1994.

[28] C. Lemaréchal and C. Sagastizábal. More than first-order developments of convex functions: primal-dual relations. *Journal of Convex Analysis*, 3(2):1–14, 1996.

[29] C. Lemaréchal and C. Sagastizábal. Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. *SIAM Journal on Optimization*, 7(2):367–385, 1997.

[30] C. Lemaréchal and C. Sagastizábal. Variable metric bundle methods: from conceptual to implementable forms. *Mathematical Programming*, 76:393–410, 1997.

[31] E. Levitin and R. Tichatschke. On smoothing of parametric minimax-functions and generalized max-functions via regularization. *Journal of Convex Analysis*, 5(2):199–220, 1998.

[32] R. Mifflin and C. Sagastizábal. $\mathcal{V}U$-decomposition derivatives for convex max-functions. In R. Tichatschke and M. Théra, editors, *Ill-posed Variational Problems and Regularization Techniques*, number 477 in Lecture Notes in Economics and Mathematical Systems, pages 167–186. Springer-Verlag Berlin Heidelberg, 1999.

[33] J.J. Moreau. Proximité et dualité dans un espace Hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[34] A.S. Nemirovski and D.B. Yudin. *Problem Complexity and Method Efficiency in Optimization.* Wiley - Interscience Series in Discrete Mathematics, 1979.

[35] J. Rey Pastor, L. Santaló, and M. Balanzat. *Geometría Analítica.* Kapelusz, Buenos Aires, 1955.

[36] G. Reinelt. Tsplib 95. Technical report, Institut für Angewandte Mathematik, Universität Heidelberg, Heidelberg, Germany, 1995. ftp://ftp.iwr.uni-heidelberg.de/pub/tsplib.

[37] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[38] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization*, 2(1):121–152, 1992.

[39] N. Shor. *Minimization methods for non-differentiable functions.* Springer-Verlag, Berlin, 1985.

[40] P. Wolfe. A method of conjugate subgradients for minimizing nondifferentiable functions. *Mathematical Programming Study*, 3:145–173, 1975.