An infeasible Predictor-Corrector Interior Point Method Applied to Image Denoising

Cecilia Pola, Claudia Sagastizábal

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

An infeasible Predictor-Corrector Interior Point method applied to Image denoising

Cecilia Pola and Claudia A. Sagastizábal

N° 3205

Juillet 1997

THÈME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

An infeasible Predictor-Corrector Interior Point method applied to Image denoising

Cecilia Pola * and Claudia A. Sagastizábal †

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Promath

Rapport de recherche n ° 3205 — Juillet 1997 — 26 pages

Abstract: Image recovery problems can be solved using optimization techniques. In this case, they often lead to the resolution of either a large scale quadratic program, or, equivalently, to a nondifferentiable minimization problem. Interior point methods are widely known for their efficiency in linear programming. Lately, they have been extended with success to the resolution of linear complementary problems, (LCP), which include convex quadratic programming. We present an infeasible predictor-corrector interior point method, in the general framework of monotone (LCP). The algorithm has polynomial complexity. We also prove it converges globally, with asymptotic quadratic rate. We apply this method to the denoising of images. In the implementation we take advantage of the underlying structure of the problem, specially its sparsity. We obtain good performances, that we assess by comparing the method with a variable-metric proximal bundle algorithm applied to the resolution of the equivalent nonsmooth problem.

Key-words: infeasible interior points, predictor-corrector methods, image denoising, kinky images, bundle methods, variable-metric proximal-point methods

(Résumé : tsvp)

* Dpto Mat., Est. y Comp., Univ. Cantabria, Spain. E-mail:pola@matesco.unican.es
† INRIA, BP 105, 78153 Le Chesnay, France and DEE- PUC-Rio. E-mail:claudia.sagastizabal@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Une méthode des points intérieurs du type prédicteur-correcteur non admissible appliquée au traitement d'images

Résumé : Le problème de récuperation d'une image bruitée peut être résolu avec des techniques issues de l'optimisation. Dans ce cas, on est amenés à résoudre soit un programme quadratique de grande taille, soit un problème de minimisation non différentiable. Les méthodes de points intérieurs sont bien connues par leur efficacité pour l'optimisation linéaire. Elles ont aussi été étendues avec succès à la résolution des problèmes de complementareité linéaire (LCP), qui incluent l'optimisation quadratique convexe. Nous présentons une méthode de points intérieurs de type prédicteur-correcteur non admissible, pour le cas général des (LCP) monotones. L'algoritme a une complexité polynomiale. Nous monstrons aussi qu'il converge globalement, avec une vitesse asymptotique d'ordre deux. Nous appliquons la méthode au traitement d'images. Pour l'implémentation, nous exploitons la structure du problème, notamment le creux des systèmes linéaires. Nous obténons de bonnes performances, que nous comparons à celles d'une méthode de faisceaux de type proximal à métrique variable, appliquée à la résolution du problème non lisse équivalent.

Mots-clé : points intérieurs non admissibles, méthodes prédicteur-correcteur, traitement d'images, contours contrastés, méthodes de faisceaux, méthodes proximales à métrique variable

1 Introduction

Consider the following image recovery problem. Let z be a noisy image described by $m := N \times N$ pixels. Its values correspond to a black-and-white unknown image x^*, corrupted by a random noise η which is uniformly distributed: $z_i = x_i^* + \eta_i$. To recover an image $\bar{x}$ as close to the original image x^* as possible, we solve an optimization problem. Let us separate pixels by rows and columns in the image-space. They define interfaces, each one of measure $1/N$, with coordinates i, j, for $i, j = 1, \ldots, N$. Then, given an arbitrary image x, reshaped as a square matrix X, its total variation in gray levels corresponds to the sum of both the vertical and the horizontal variations:

$$TV(x) = \sum_{i=1}^{N} \sum_{j=1}^{N-1} \frac{1}{N} |X_{i,j} - X_{i,j+1}| + \sum_{j=1}^{N} \sum_{i=1}^{N-1} \frac{1}{N} |X_{i,j} - X_{i+1,j}|. \tag{1}$$

Total variations can be used in the discretization of some optimization problems that are regularized by bounded-variation seminorms, see for example [2]. This approach leads to the resolution of an optimization problem

$$\min_{x \in \mathbb{R}^m} \frac{\alpha}{2m} \sum_{i=1}^{m} \|x_i - z_i\|_2^2 + TV(x)\,, \tag{2}$$

where α is a positive parameter.

We can rewrite (2) by expressing $TV(\cdot)$ from (1) as the following norm:

$$TV(x) \quad = \quad \|Ax\|_1\,, \quad \text{with} \quad A = \frac{1}{n} \begin{pmatrix} I & \otimes & M \\ M & \otimes & I \end{pmatrix}. \tag{3}$$

Here, $\otimes$ denotes the Kronecker product, I is the $N \times N$ identity matrix and M is the $(N-1) \times N$ matrix given by

$$M = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 & -1 \end{pmatrix}.$$

With this notation, (2) becomes

$$\min_{x \in \mathbb{R}^m} \frac{\alpha}{2m} x^T x + \frac{\alpha}{m} z^T x + \|Ax\|_1\,. \tag{4}$$

Clearly, (4) is a nonsmooth optimization problem: the objective has kinks at any x such that $(Ax)_i = 0$, for some index i. Therefore, it can be solved using bundle techniques. An

alternative is to formulate an equivalent quadratic program, but with higher dimension:

$$\left\{ \begin{array}{c} \min \\ (x,u,v) \in \mathbb{R}^{m+r+r} \\ Ax - u + v = 0 \\ u, v \geq 0\,, \end{array} \right. \quad \frac{\alpha}{2m} x^T x + \frac{\alpha}{m} z^T x + e^T (u+v) \tag{5}$$

where we noted $r = 2N(N-1)$ and $e = (1, \ldots, 1)^T \in \mathbb{R}^r$. Problems (5) and (4) are equivalent: if $(\bar{x}, \bar{u}, \bar{v})$ is solution of (5), then $\bar{x}$ is solution of (4). Reciprocally, given any solution $\bar{x}$ of (4), the triplet $(\bar{x}, \bar{u} := \max\{A\bar{x}, 0\}, \bar{v} := \max\{-A\bar{x}, 0\})$ solves (5).

Therefore, to recover the image, we can solve (4) or (5). We adopt the later formulation, that can be embedded in the general framework of monotone linear complementarity problems, (LCP). More precisely, we use an interior point method that we present in the sequel.

The paper is organized as follows. In the next section we present an infeasible predictor-corrector interior point algorithm, (PCIP), for (LCP). We study convergence of the algorithm in sections 3 and 4. We show first that (PCIP) has polynomial complexity and that it is globally convergent. In § 4 we prove that (PCIP) converges with asymptotic quadratic rate, provided there is strict complementarity on (LCP). The remaining sections are devoted to numerical considerations. Section 5 deals with computational aspects of (PCIP), exploiting the underlying structure in (5), in particular its sparsity. We outline in § 6 the proximal-quasi-Newton bundle method that we apply to the resolution of (4), with comparing purposes. Finally, we assess our approach with some numerical results, reported in Section 7.

We adopt the following convention. Given a vector $x \in \mathbb{R}^m$, vectorial inequalities like $x \geq 0$ are taken componentwise: $x_i \geq 0$ for any $i = 1, 2, \cdots, m$. We denote by $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ the nonnegative and positive orthants:

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geq 0\} \qquad \mathbb{R}^n_{++} = \{x \in \mathbb{R}^n : x > 0\}\,.$$

When refering to a norm other than the Euclidean norm, we add a subindex, like $\|\cdot\|_1$ or $\|\cdot\|_\infty$. Finally, usual arithmetic operations for real numbers are applied here to vectors componentwise: $uv = (u_i v_i)_i$, $u/v = (u_i/v_i)_i$, $u^{-1} = (1/u_i)_i$, and so on.

2 An infeasible predictor corrector algorithm

Problem (5) can be considered a particular case of the following (monotone horizontal) linear complementarity problem

$$\text{Find } (y,s) \in \mathbb{R}^n \times \mathbb{R}^n \text{ such that} \qquad \left\{ \begin{array}{rcl} ys & = & 0 \\ Qy + Rs & = & h \\ y\,, s & \geq & 0\,, \end{array} \right. \tag{LCP}$$

where $h \in \mathbb{R}^n$, and $Q, R \in \mathbb{R}^{n\times n}$ satisfy the *monotonicity* equation

$$Qd_y + Rd_s = 0 \text{ implies } d_y^T d_s \geq 0\,, \quad \text{for all } d_y, d_s \in \mathbb{R}^n. \tag{6}$$

When particularizing to problem (5), y gathers the unknows: positive and negative part of x, as well as u and v. As for s, it denotes the corresponding dual variables associated to the positivity constraints in (5).

Associated to (LCP) there is the set of strictly complementary solutions:

$$\mathcal{S}^0 = \{(y,s) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \,:\, Qy + Rs = h\,,\, ys = 0 \text{ and } y + s > 0\}\,. \tag{7}$$

For proving quadratic convergence in § 4 we will assume the existence of a strictly complementary solution, i.e., $\mathcal{S}^0 \neq \emptyset$.

Before describing how infeasible predictor-corrector methods work, some definitions and notation are needed. We adopt essentially the notation in [4].

Definition 1

– Given $(y^0, s^0, \mu_0) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}_{++}$, take

$$b = (h - Qy^0 - Rs^0)/\mu_0\,.$$

Then the *infeasible central path pinned on* b is defined by the triplets (y, s, μ) such that

$$\begin{array}{rcl} ys & = & \mu e \\ Qy + Rs & = & h - \mu b\,. \end{array} \tag{8}$$

– We denote by $\mathcal{F}^b$ the set of (positive) points such that the second constraint in (8) holds:

$$\mathcal{F}^b = \{(y,s,\mu) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}_{++} \,:\, Qy + Rs = h - \mu b\}\,.$$

Observe that $(y^0, s^0, \mu_0) \in \mathcal{F}^b$. Moreover, if $y^0 s^0 = \mu_0 e$, then it lies in the infeasible central path pinned on b. Also, note that $b = 0$ means that (y^0, s^0, μ) is feasible in (LCP) for any $\mu > 0$.

– The iterates in our algorithm are contained in a very special neighbourhood, defined as follows. For some constant $\nu \in (0,1)$, consider first the set

$$\mathcal{T}_\nu = \{\xi \in \mathbb{R}^n \,:\, \nu e \leq \xi \leq \nu^{-1} e\}\,;$$

then the *safety* neighbourhood is

$$\mathcal{N}^b_\nu = \{(y,s,\mu) \in \mathcal{F}^b \,:\, \frac{ys}{\mu} \in \mathcal{T}_\nu \text{ and } \mu \leq \mu_0\}\,.$$

– To measure how far from the border of $\mathcal{T}_\nu$ a point is, we use the function

$$\mathcal{D}(\xi) = \min_{i=1,\ldots,n} \left(|\xi_i - \nu|, |\xi - \nu^{-1}|\right) = \min\left(|\xi - \nu e|, |\xi - \nu^{-1} e|\right)\,.$$

□

Let us now explain one step of the algorithmic pattern. At the current iteration, a point $(y, s, \mu) \in \mathcal{N}_\nu^b$ is available.

- Consider (8), where μ has been replaced by $\gamma\mu$, for some $\gamma \in [0, 1]$. An approximate solution can be found by making a Newton's step from (y, s) to $(y + d_y, s + d_s)$:

$$\begin{array}{rcl} s d_y + y d_s & = & -ys + \gamma\mu e \\ Q d_y + R d_s & = & (1-\gamma)\mu b\,. \end{array} \tag{9}$$

- Under the monotonicity assumption (6), the system has a unique solution $(d_y^\sharp, d_s^\sharp)$, which can be interpreted as the convex sum of an *affine-scaling* step (prediction) and a *centering* step (correction):

  Rewrite the first equation in (9) to obtain

$$\begin{array}{rcl} s d_y + y d_s & = & (1-\gamma)(-ys) + \gamma(-ys + \mu e) \\ Q d_y + R d_s & = & (1-\gamma)\mu b\,. \end{array} \tag{9}_\gamma$$

  Then

$$(d_y^\sharp, d_s^\sharp) = (1-\gamma)(d_y^a, d_s^a) + \gamma(d_y^c, d_s^c)\,,$$

  where (d_y^a, d_s^a) is the solution of $(9)_{\gamma=0}$ and (d_y^c, d_s^c) solves $(9)_{\gamma=1}$.

- Finally, choose an stepsize $\theta \in (0, 1]$ and $\gamma \in [0, 1]$ such that the updates

$$\begin{array}{rcll} y^\sharp & = & y + \theta d_y^\sharp & > \ 0 \\ s^\sharp & = & s + \theta d_s^\sharp & > \ 0 \\ \mu_\sharp & = & (1 - \theta + \theta\gamma)\mu & \leq \ \mu \leq \mu_0\,, \end{array} \tag{10}$$

  remain in the neighbourhood $\mathcal{N}_\nu^b$. Replace (y, s, μ) with the values $(y^\sharp, s^\sharp, \mu_\sharp)$. □

We are now in position to give the algorithm with full details.

Predictor-Corrector Interior Point algorithm (PCIP)

STEP 0 Given an initial $(y^0, s^0, \mu_0) \in \mathcal{N}_\nu^b$, choose parameters $\beta \in (0, 1/2]$, $\nu \in (0, 0.1]$ and $\mu_\infty > 0$. Set to zero the counter: $k = 0$.

STEP 1 For $y = y^k$, $s = s^k$, $\mu = \mu_k$:

–Compute (d_y^c, d_s^c), the solution of $(9)_{\gamma=1}$. **(correction)**

–Compute $\theta^c \in (0, 1]$ such that **(centering stepsize)**

$$\theta^c = \min\left\{1, \frac{(1-\nu)}{2}\frac{\mu}{\|d_y^c d_s^c\|_\infty}\right\}\,. \tag{11}$$

STEP 2 For $y = y + \theta^c d_y^c,\ s = s + \theta^c d_s^c$:

–Compute (d_y^a, d_s^a), the solution of $(9)_{\gamma=0}$. **(prediction)**

STEP 3 **(affine stepsize)**

–Compute $\hat{\theta}^a = \dfrac{-1+\sqrt{1+4\eta}}{2\eta}$, where $\eta = \dfrac{\|d_y^a d_s^a\|_\infty}{\mu}\dfrac{\sqrt{2}n}{\nu^3}$.

–Set $\bar{\theta}^a = 1/\max\{\max\{[-d_y^a/y; -d_s^a/s]\}, 1\}$.

–Compute $\bar{r} = \min\{r \in \mathbb{N} : \mathcal{D}(\dfrac{(y+\beta^r\bar{\theta}^a d_y^a)(s+\beta^r\bar{\theta}^a d_s^a)}{\mu}) > 0\}$.

Take $\theta^a = \max\{\beta^{\bar{r}}\bar{\theta}^a, \hat{\theta}^a\}$.

STEP 4 **(update)**

–Define $y^{k+1} = y + \theta^a d_y^a,\ s^{k+1} = s + \theta^a d_s^a,\ \mu_{k+1} = (1-\theta^a)\mu_k$.

–Update $k = k+1$. If $\mu_k > \mu_\infty$, then loop to 1, else stop.

□

Some comments are now in order. Relative to STEP 3, a simpler expression, like

STEP 3' Let θ^a be the largest stepsize in $(0,1)$ such that

$$(y + \theta^a d_y^a, s + \theta^a d_s^a, (1-\theta^a)\mu) \in \mathcal{N}_\nu^b\,,$$

would still ensure our convergence results. However, we prefer to give a less abstract criterion for θ^a, which can be implemented and are still suitable for proving convergence.

We will see in § 4 that $\hat{\theta}^a$ plays a crucial role for having quadratic convergence, while $\beta^{\bar{r}}\bar{\theta}^a$ is used for getting polynomial complexity. The particular choices of θ^c and $\hat{\theta}^a$ will be clarified when proving Lemma 2 and Lemma 6 respectively.

In the following sections we address convergence issues of (PCIP), considering first its polynomial complexity, as well as its global convergence. Finally, we show that (PCIP) converges asymptotically with quadratic rate.

3 Polynomial complexity and global convergence

To showing that (PCIP) has a complexity of order $O(n\sqrt{n}L)$, we split our analysis considering the corrector step on one side, and the predictor step on the other. Both steps are obtained from solving $(9)\gamma$, for $\gamma = 1$ and $\gamma = 0$ respectively. Thus, here we find it convenient to scale $(9)\gamma$ as follows: given

$$\phi = \sqrt{\frac{ys}{\mu}};\ \delta = \sqrt{\frac{\mu y}{s}};\ \bar{d}_y = \delta^{-1}d_y;\ \bar{d}_s = \delta\mu^{-1}d_s;\ \bar{Q} = QI\delta;\ \bar{R} = \mu RI\delta^{-1}\,, \qquad (12)$$

multiply the first equation in (9) by $1/\sqrt{\mu ys}$, to obtain

$$\sqrt{\frac{s}{\mu y}}d_y + \frac{1}{\mu}\sqrt{\frac{\mu y}{s}}d_s \;= -\sqrt{\frac{ys}{\mu}} + \gamma\sqrt{\frac{\mu}{ys}}\,.$$

Denoting with an upper bar those variables which are scaled, we have that $(\bar{d}^\sharp_y, \bar{d}^\sharp_s) = (\delta^{-1}d^\sharp{}_y, \delta\mu^{-1}d^\sharp{}_s)$ is the solutoin of

$$\left\{\begin{array}{lcll} \bar{d}_y + \bar{d}_s & = & -\phi + \gamma\phi^{-1} & =: \; f^\sharp \\ \bar{Q}\bar{d}_y + \bar{R}\bar{d}_s & = & (1-\gamma)\mu b & =: \; b^\sharp\,. \end{array}\right. \tag{13}$$

Also, combining (9) and (10) we have that

$$\frac{y^\sharp s^\sharp}{\mu_\sharp} = \frac{1-\theta}{1-\theta+\theta\gamma}\left(\frac{ys}{\mu} - e\right) + e + \frac{\theta^2}{1-\theta+\theta\gamma}\frac{d^\sharp{}_y d^\sharp{}_s}{\mu}\,, \tag{14}$$

where $(y^\sharp, s^\sharp) = (y,s) + \theta(d^\sharp_y, d^\sharp_s)$ and $\mu_\sharp = (1-\theta+\theta\gamma)\mu$ are those defined in (10).

Analysis of the Corrector step After performing STEP 1, the corrector step has been made, and we have computed (d^c_y, d^c_s) and θ^c. We prove that the new point $(y^c, s^c) = (y + \theta^c d^c_y, s + \theta^c d^c_s)$ is in the safety neighbourhood $\mathcal{N}^b_\nu$.

Lemma 2 *Consider* STEP 1 *in (PCIP). Let $(y,s,\mu) \in \mathcal{N}^b_\nu$ and (y^c, s^c, μ) denote respectively the current and "corrector" iterates. Then the following holds:*

(i) The centering direction satisfies

$$\frac{\|d^c_y d^c_s\|_\infty}{\mu} \;\leq\; \frac{n(1-\nu)^2}{2\sqrt{2}\nu^3}\,.$$

(ii) For all $\nu \in (0, 1/2]$, the triplet $(y^c, s^c, \mu) \in \mathcal{N}^b_\nu$. Moreover,

$$\mathcal{D}(\frac{y^c s^c}{\mu}) \geq \frac{\nu^3\sqrt{2}}{2n}\,. \tag{15}$$

Proof. Let us prove (i). We have that $(y,s,\mu) \in \mathcal{N}^b_\nu$. Then it suffices to combine Mizuno's Lemma ([6, Lemma 1]) with (13) to obtain the following chain of inequalities:

$$\frac{\|d^c_y d^c_s\|_\infty}{\mu} \leq \frac{\|d^c_y d^c_s\|}{\mu} \leq \frac{1}{\nu\sqrt{8}}\left\|\frac{ys}{\mu} - e\right\|^2 \leq \frac{n}{\nu\sqrt{8}}\left\|\frac{ys}{\mu} - e\right\|^2_\infty \leq \frac{n}{2\sqrt{2}\nu}\left(\frac{1}{\nu} - 1\right)^2\,.$$

For proving (ii), we show first that the element $\xi(\theta) = \dfrac{(y+\theta d^c_y)(s+\theta d^c_s)}{\mu}$ belongs to $\mathcal{T}_\nu$, for all $\theta \in [0, \theta^c]$. For this, use (14) with $\gamma = 1$ to write

$$\begin{aligned}\xi(\theta) &= (1-\theta)\left(\frac{ys}{\mu}-e\right)+e+\theta^2\frac{d_y^c d_s^c}{\mu}\\ &= (1-\theta)\frac{ys}{\mu}+\theta e+\theta^2\frac{d_y^c d_s^c}{\mu}\,,\end{aligned}$$

so that

$$\mathcal{D}(\xi(\theta)) \geq \mathcal{D}((1-\theta)\frac{ys}{\mu}+\theta e)-\theta^2\frac{\|d_y^c d_s^c\|_\infty}{\mu}\,. \tag{16}$$

Let us bound from above the first righthand side term in (16). Because $(y,s,\mu)\in\mathcal{N}_\nu^b$, $ys/\mu\in\mathcal{T}_\nu$. Also, $\theta\in(0,1]$ and $\nu<1$. Altogether,

$$\mathcal{D}((1-\theta)\frac{ys}{\mu}+\theta e)\geq\theta(1-\nu)\,.$$

Then (16) yields

$$\mathcal{D}(\xi(\theta))\geq\theta(1-\nu)-\theta^2\frac{\|d_y^c d_s^c\|_\infty}{\mu}\,. \tag{17}$$

As a function of θ, the righthand side $\varphi:[0,\theta^c]\longrightarrow\mathbb{R}$ defined by

$$\varphi(\theta)=\theta(1-\nu)-\theta^2\frac{\|d_y^c d_s^c\|_\infty}{\mu}$$

is increasing; furthermore, $\varphi(0)=0$. Together with (17), we obtain that

$$\mathcal{D}(\xi(\theta))\geq 0\,,\quad \text{for all } \theta\in[0,\theta^c]\,. \tag{18}$$

To prove that (y^c,s^c,μ) is in the safety neighbourhood, it only remains to show that the triplet is in $\mathcal{F}^b$. We just need to verify the positivity constraints. Combine the definition of $\mathcal{D}(\cdot)$ with (18) to write

$$0<\nu<\frac{(y+\theta d_y^c)(s+\theta d_s^c)}{\mu}\,,\quad \text{for all } \theta\in[0,\theta^c]\,.$$

In particular, because $y,s>0$ and $\mu>0$,

$$0<y+\theta^c d_y^c,\ 0<s+\theta^c d_s^c\,.$$

Altogether, $(y^c,s^c,\mu)\in\mathcal{N}_\nu^b$.

Finally, we prove (15). If $\theta^c=1$, because of (11), we have that $\dfrac{\|d_y^c d_s^c\|_\infty}{\mu}\leq\dfrac{(1-\nu)}{2}$. Thus, $\mathcal{D}(\xi(\theta^c))\geq\dfrac{1-\nu}{2}$ and we are done. Else, if $\theta^c<1$, from (11) we have that

$$\frac{\|d_y^c d_s^c\|_\infty}{\mu}\geq\frac{1-\nu}{2}\,.$$

Then, in (17) we get $\mathcal{D}(\xi(\theta^c)) \geq \dfrac{(1-\nu)^2}{4}\dfrac{\mu}{\|d_y^c d_s^c\|_\infty}$ and the conclusion follows. □

Observe that the value $\dfrac{(1-\nu)}{2}\dfrac{\mu}{\|d_y^c d_s^c\|_\infty}$, introduced in STEP 1 of (PCIP) for defining θ^c, is the maximizer of the function φ used in the proof above. This explains our choice of the centralizing stepsize θ^c in (PCIP): it drives (y^c, s^c, μ) "as far as possible" from the border of $\mathcal{T}_\nu$.

Analysis of the Predictor step Here, STEP 3 has been made and we have $(y^{k+1}, s^{k+1}) = (y + \theta^a d_y^a, s + \theta^a d_s^a)$. To show that this new point belongs to $\mathcal{N}_\nu^b$, some technical results are needed and the proofs are slightly more involved. In particular, we need to introduce the following concept, see [4]:

Definition 3 Let (y^*, s^*) be a solution of (LCP):

$$Qy^* + Rs^* = h \quad \text{and} \quad y^* s^* = 0\,.$$

We say that the point $(y^0, s^0) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+$ *dominates* (y^*, s^*) if $y^0 > y^*$ and $s^0 > s^*$. □

We will also use two results that we state here without proof:

Lemma 4 (Corollary 1 in [7]) *Let $(\bar{d}_y^\sharp, \bar{d}_s^\sharp)$ solve (13) and let $\hat{y}, \hat{s}$ be such that $Q\hat{y} + R\hat{s} = b^\sharp$. Then*

$$\|\bar{d}_y^\sharp\| \leq \|f^\sharp\| + \|\hat{y}\| + \|\hat{s}\| \qquad \text{and} \qquad \|\bar{d}_s^\sharp\| \leq \|f^\sharp\| + \|\hat{y}\| + \|\hat{s}\|\,.$$

□

Lemma 5 (Lemma 3.2 in [4]) *Let (y^0, s^0, μ_0) and (y, s, μ) be two elements of $\mathcal{N}_\nu^b$ such that (y^0, s^0) dominates a solution (y^*, s^*) of (LCP). Then*

$$y^T s^0 + s^T y^0 \leq 4\mu_0 \frac{n}{\nu}\,.$$

Moreover,

$$Q\tilde{y} + R\tilde{s} = b \qquad \text{and} \qquad \|\delta^{-1}\tilde{y}\| + \|\delta\mu^{-1}\tilde{s}\| \leq \frac{4n}{\nu\sqrt{\nu}\mu}\,,$$

for $(\tilde{y}, \tilde{s}) = (y^* - y^0, s^* - s^0)/\mu^0$ and δ from (12).

□

Now we find a lower bound for the affine step θ^a.

Lemma 6 *Consider* STEPS 2-3 *in (PCIP). Let $(y, s, \mu) \in \mathcal{N}_\nu^b$ and $(y^{k+1}, s^{k+1}, \mu_{k+1})$ denote respectively the current and "predictor" iterates. Assume that (y^0, s^0) dominates a solution (y^*, s^*) of (LCP). If $\mu \leq \mu_0$, then the following holds:*

(i) The affine direction satisfies

$$\frac{\|d_y^a d_s^a\|}{\mu} \leq \frac{25n^2}{\nu^3}\,.$$

(ii) If $\mathcal{D}(\frac{ys}{\mu}) \geq \frac{\nu^3\sqrt{2}}{2n}$, *then*

- *for all* $\theta \in [0, \hat{\theta}^a]$,

$$\mathcal{D}(\frac{(y+\theta d_y^a)(s+\theta d_s^a)}{(1-\theta)\mu}) \geq \frac{\nu^3\sqrt{2}}{2n} - \frac{\theta^2}{1-\theta}\frac{\|d_y^a d_s^a\|_\infty}{\mu}\,. \tag{19}$$

- *The triplet* $(y^{k+1}, s^{k+1}, \mu_{k+1}) \in \mathcal{N}_\nu^b$.
- *Finally,* $\theta^a \geq \frac{\nu^3\sqrt{2}}{20n\sqrt{n}}$.

Proof. Let us prove (i). Consider (13) written with $\gamma = 0$. From (12) we have that $\|\phi^{-1}\|_\infty \leq 1/\sqrt{\nu}$ and $\|\phi\| \leq \sqrt{n/\nu}$, as well as $(\bar{d}_y^1, \bar{d}_s^1) = (\delta^{-1}d_y^a, \delta\mu^{1-}d_s^a)$. This, together with Lemmas 4-5 yields

$$\|\delta^{-1}d_y^a\| \leq \|-\phi\| + \mu\left(\|\delta^{-1}\tilde{y}\| + \|\delta\mu^{-1}\tilde{s}\|\right) \leq \frac{\sqrt{n}}{\sqrt{\nu}} + 4\frac{n}{\nu\sqrt{\nu}} \leq \frac{5n}{\nu\sqrt{\nu}}\,,$$

and the same estimate holds for $\delta\mu^{-1}d_s^a$. Altogether, because

$$\frac{\|d_y^a d_s^a\|}{\mu} = \|\delta^{-1}d_y^a\delta\mu^{-1}d_s^a\| \leq \|\delta^{-1}d_y^a\|\|\delta\mu^{-1}d_s^a\|\,,$$

the result follows.

For proving (19), define the function $\xi(\theta) = \dfrac{(y+\theta d_y^a)(s+\theta d_s^a)}{(1-\theta)\mu}$, for all $\theta \in [0,1)$ and use (14) with $\gamma = 0$:

$$\xi(\theta) \quad = \quad \frac{ys}{\mu} + \frac{\theta^2}{1-\theta}\frac{d_y^a d_s^a}{\mu}\,,$$

so that

$$\mathcal{D}(\xi(\theta)) \geq \mathcal{D}(\frac{ys}{\mu}) - \frac{\theta^2}{\{1-\theta}\frac{\|d_y^a d_s^a\|_\infty}{\mu}\,.$$

Because of the assumption in (ii), (19) follows.

As a function of θ, the righthand side in (19), $\varphi : [0, \hat{\theta}^a] \longrightarrow \mathbb{R}$ defined by

$$\varphi(\theta) = \frac{\nu^3\sqrt{2}}{2n} - \frac{\theta^2}{1-\theta}\frac{\|d_y^a d_s^a\|_\infty}{\mu}$$

is decreasing; furthermore, $\varphi(\hat{\theta}^a) = 0$. Together with (19), we conclude that the triplet $(y+\theta d_y^a, s+\theta d_s^a, (1-\theta)\mu)$ is in the safety neighbourhood, for all $\theta \in [0, \hat{\theta}^a]$.

On the other side, the triplet $(y+\beta^{\bar r}\bar\theta^a d_y^a, s+\beta^{\bar r}\bar\theta^a d_s^a, (1-\beta^{\bar r}\bar\theta^a)\mu)$ is also an element of $\mathcal{N}_\nu^b$. From the definition of θ^a in STEP 3, it follows that $(y^{k+1}, s^{k+1}, \mu_{k+1}=(1-\theta^a)\mu) \in \mathcal{N}_\nu^b$.

Finally, use again (19) together with (i) to write

$$\mathcal{D}(\xi(\theta)) \geq \frac{\nu^3\sqrt{2}}{2n} - 2\theta^2\frac{25n^2}{\nu^3}\,,$$

so that

$$(y+\theta d_y^a, s+\theta d_s^a, (1-\theta)\mu) \in \mathcal{N}_\nu^b\,, \quad \text{for all } \theta \in (0,\bar\theta]\,,$$

where $\bar\theta = \dfrac{\nu^3}{10}\sqrt{\sqrt{2}}\dfrac{1}{n\sqrt{n}}$. Altogether, recalling again the choice of θ^a, we have $\theta^a > \bar\theta/2$ and the proof is finished. □

Observe that the value $\hat\theta^a$, introduced in STEP 3 of (PCIP) for defining θ^a, is now a zero of the function φ used in the proof above. This guarantees that $(y+\hat\theta^a d_y^a, s+\hat\theta^a d_s^a, (1-\hat\theta^a)\mu)$ remains inside of $\mathcal{N}_\nu^b$.

The assumption in (ii) may seem artificial. Note, however, that at this stage $(y,s,\mu) = (y^c, s^c, \mu)$. Thanks to Lemma 2$(i)$, the assumption is naturally satisfied.

Now we can state our first result of convergence of (PCIP), relative to polynomial complexity and global convergence:

Theorem 7 *Assume that (y^0, s^0) dominates a solution of (LCP) and consider the sequence $\{(y^k, s^k, \mu_k)\}$ generated by (PCIP). Then the following holds:*

- *(i) If $\mu_\infty > 0$ then (PCIP) has polynomial complexity: there exists $\bar k = O(n\sqrt{n}L)$ such that $\mu_{\bar k} \leq \mu_\infty$ in* STEP *4, where we have set $L = \log_2(\mu_0/\mu_\infty)$.*
- *(ii) If $\mu_\infty = 0$ then both the gap $y^k s^k$ and the residuals $h - Qy^k - Rs^k$ tend to 0 when k goes to infinity.*

Proof. Let us prove (i). Because of Lemma 2(ii), $\mathcal{D}(\dfrac{y^c s^c}{\mu}) \geq \dfrac{\nu^3\sqrt{2}}{2n}$ and the assumption in Lemma 6(ii) holds; hence, the affine stepsize is bounded from below:

$$\theta^a \geq \frac{\nu^3\sqrt{2}}{20n\sqrt{n}}\,.$$

Now, combine this inequality with the update of μ in STEP 4 to obtain

$$\mu_k = (1-\theta^a)\mu_{k-1} \leq \left(1 - \frac{\nu^3\sqrt{2}}{20n\sqrt{n}}\right)^k \mu_0\,. \tag{20}$$

Taking logarithms, and using that $|\log_2(1-t)| \geq t$, for $t = \dfrac{\nu^3\sqrt{2}}{20n\sqrt{n}}$, the result follows.

For proving (ii), note that $\mu_\infty = 0$ implies that $k \to \infty$. Together with (20), this means that $\mu_k \to 0$. Since the triplet (y^k, s^k, μ_k) is in $\mathcal{N}_\nu^b$, we have the bounds

$$\nu\mu_k e \leq y^k s^k \leq \nu^{-1}\mu_k e\,,$$

so that in the limit $y^k s^k \to 0$.

Consider now the residuals $r^k := h - Qy^k - Rs^k$. We claim that $r^k = \mu_k b$. Indeed, when $k = 1$, recalling the definition of y^1 and s^1 and using (9) as well as the expression for b in Definition 1, we have that

$$\begin{aligned} r^1 &= h - Q(y^0 + \theta_0^c d_{y^0}^c + \theta_0^a d_{y^0}^a) - R(s^0 + \theta_0^c d_{s^0}^c + \theta_0^a d_{s^0}^a) \\ &= (h - Qy^0 - Rs^0) - \theta_0^c(Qd_{y^0}^c + Rd_{s^0}^c) - \theta_0^a(Qd_{y^0}^a + Rd_{s^0}^a) \\ &= r_0 - 0 - \theta_0^a\mu_0 b \\ &= \mu_0 b - \theta_0^a\mu_0 b \\ &= \mu_1 b\,. \end{aligned}$$

For the inductive step, suppose that $r^k = \mu_k b$. Then, because $r^{k+1} = h - Qy^{k+1} - Rs^{k+1} = r^k - \theta_k^a\mu_k b = \mu_{k+1}b$, the claim is proved. □

4 Asymptotic rate of convergence

Throughout this section we assume that there exists a strictly complementary solution for (LCP): $\mathcal{S}^0 \neq \emptyset$. Under this assumption, a unique partition of the index set $\{1, \ldots, n\}$ can be defined as follows:

$$B \cup N = \{1, 2, \cdots, n\} \quad \text{with} \quad B \cap N = \emptyset\,,$$

such that

$$\text{if } (y, s) \in \mathcal{S}^0 \quad \text{then} \begin{cases} y_B > 0\,, & s_B = 0 \\ y_N = 0\,, & s_N > 0\,. \end{cases}$$

This partition can be used to rearrange variables (see [3]):

$$y \leftarrow (y_B, s_N) \quad \text{and} \quad s \leftarrow (y_N, s_B)\,,$$

so that the new y gathers only *large* variables, while s has the *small* ones. The naming is related to the following relation (Lemma 4.1 in [1]), useful when studying asymptotic behaviour:

$$\text{For any rearranged } (y, s, \mu) \in \mathcal{N}_\nu^b \text{ it holds that} \quad y = O(1) \quad \text{and} \quad s = O(\mu)\,, \tag{21}$$

Note that (PCIP) is invariant with respect to the index permutation, whose advantage is just to simplify the convergence analysis by assuming that $N = \emptyset$.

After rearranging, we scale y and s like in (12) and (13) from § 3. In this way, we can express the (rearranged and scaled) affine direction in terms of $O(\mu)$:

Lemma 8 *Assume that (LCP) is such that $\mathcal{S}^0 \neq \emptyset$ and consider one iteration of (PCIP). The affine direction satisfies*

$$\frac{\|d_y^a d_s^a\|_\infty}{\mu} = O(\mu)\,.$$

Proof. We have that $(y, s, \mu) \in \mathcal{N}_\nu^b$. The strict complementary implies that there exists a point (y', s') such that $Qy' + Rs' = b$. Then Lemma 3.5 in [4] applies, written with $(\bar{u}, \bar{v}) = (\bar{d}_y^a, \bar{d}_s^a)$ and $\gamma = \eta = 0$:

$$\bar{d}_y^a = O(\mu) \quad \text{and} \quad \bar{d}_s^a = -\phi + O(\mu)\,,$$

where ϕ was defined in (12). Using (12), the expression above leads to

$$d_y^a d_s^a = \mu \bar{d}_y^a \bar{d}_s^a = O(\mu)\,(\phi + O(\mu))\ .$$

If we showed that $\phi = O(1)$, the proof would be finished. To see this, just use again (12): $\phi = \sqrt{\dfrac{ys}{\mu}}$, together with (21), the conclusion follows. □

With this last Lemma, the quadratic rate of convergence is straightforward. For the sake of completeness, we gather here all the convergence results proved for (PCIP).

Theorem 9 *Assume that (y^0, s^0) dominates a solution of (LCP) and consider the sequence $\{(y^k, s^k, \mu_k)\}$ generated by (PCIP). Then the following holds:*

(i) If $\mu_\infty > 0$ then (PCIP) has polynomial complexity: there exists $\bar{k} = O(n\sqrt{n}L)$ such that $\mu_{\bar{k}} \leq \mu_\infty$ in Step 4*, where we have set $L = \log_2(\mu_0/\mu_\infty)$.*

(ii) If $\mu_\infty = 0$ then both the gap $y^k s^k$ and the residuals $h - Qy^k - Rs^k$ tend to 0 when k goes to infinity.

(iii) If $\mu_\infty = 0$ and, in addition, $\mathcal{S}^0 \neq \emptyset$, then $\mu_k \longrightarrow 0$ Q-quadratically.

Proof. (i) and (ii) are Theorem 7. Let us prove (iii). In Step 4, we have the update $\mu^{k+1} = (1 - \theta^a)\mu^k$. Then, we only need to show that $1 - \theta^a = O(\mu_k)$.

From (19) and the definition of $\hat{\theta}^a$, we have that

$$\mathcal{D}\Big(\frac{(y + \theta d_y^a)(s + \theta d_s^a)}{(1-\theta)\mu}\Big) \geq 0 \quad \text{for all } \theta \in (0, \hat{\theta}^a]\,.$$

Recall now the definitions of η and $\hat{\theta}^a$ in Step 3 of (PCIP); the following relation holds: $(1 - \hat{\theta}^a) = (\hat{\theta}^a)^2 \dfrac{\|d_y^a d_s^a\|_\infty}{\mu_k} \dfrac{\sqrt{2}n}{\nu^3} = O(\mu_k)$. The conclusion follows, because $\theta^a \geq \hat{\theta}^a$. □

5 Implementing (PCIP)

Here, rather than treating a general (LCP), we address some practical issues taking advantage of the particular structure of (5),

Consider the constraints in (5): associated to the equality constraint there is the vector $\lambda \in \mathbb{R}^r$, while positivity constraints on u and v have (nonnegative) multipliers s_u, s_v, both in $\mathbb{R}^r$. Therefore, the optimality conditions are

$$\left\{\begin{array}{rcl} us_u & = & 0 \\ vs_v & = & 0 \\ -\frac{\alpha}{m}x + A^T\lambda & = & p \\ -\lambda + s_u & = & e \\ \lambda + s_v & = & e \\ Ax - u + v & = & 0 \\ u\,,\, v\,,\, s_u\,,\, s_v & \geq & 0\,, \end{array}\right. \tag{22}$$

where we set $p = \dfrac{\alpha}{m}z$, with z the initial noisy image.

Take initial values x^0, u^0, v^0, s_u^0, s_v^0, λ^0 and μ_0. At the current iteration, given a target value of μ, the system (9) corresponding to a Newton step from $(x, u, v, s_u, s_v, \lambda)$ to $(x + d_x, u + d_u, v + d_v, s_u + d_{s_u}, s_v + d_{s_v}, \lambda + d_\lambda)$ is

$$\gamma\left\{\begin{array}{rcl} ud_{s_u} + s_u d_u & = & -us_u + \gamma\mu e \\ vd_{s_v} + s_v d_v & = & -vs_v + \gamma\mu e \\ -\frac{\alpha}{m}d_x + A^T d_\lambda & = & \sigma_1 \\ -d_\lambda + d_{s_u} & = & \sigma_2 \\ d_\lambda + d_{s_v} & = & \sigma_3 \\ Ad_x - d_u + d_v & = & \rho, \end{array}\right. \tag{23}$$

where we have set

$$\sigma_1 = (1-\gamma)(p + \frac{\alpha}{m}x^0 - A^T\lambda^0)\frac{\mu}{\mu_0}\,,\quad \sigma_2 = (1-\gamma)(e + \lambda^0 - s_u^0)\frac{\mu}{\mu_0}\,,$$
$$\sigma_3 = (1-\gamma)(e - \lambda^0 - s_v^0)\frac{\mu}{\mu_0} \text{ and } \quad \rho = (1-\gamma)(-Ax^0 + u^0 - v^0)\frac{\mu}{\mu_0}\,.$$

Each iteration in (PCIP) requires the resolution of two systems like $(23)_\gamma$, one for the correction step ($\gamma = 0$) and other for the prediction ($\gamma = 1$). To alleviate the computational burden;, it is important to use an efficient solver. Note that $(23)_\gamma$ is a linear system which is large (for N pixels, it has dimension $11N^2 - 10N$) and indefinite, though sparse. By

properly rearranging rows and columns it can be made symmetric:

$$\begin{pmatrix} 0 & 0 & 0 & I & I & 0 \\ 0 & 0 & 0 & -I & 0 & I \\ 0 & 0 & -\frac{\alpha}{m}I & A^T & 0 & 0 \\ I & -I & A & 0 & 0 & 0 \\ I & 0 & 0 & 0 & s_v^{-1}v & 0 \\ 0 & I & 0 & 0 & 0 & s_u^{-1}u \end{pmatrix} \begin{pmatrix} d_v \\ d_u \\ d_x \\ d_\lambda \\ d_{s_v} \\ d_{s_u} \end{pmatrix} = \begin{pmatrix} \sigma_3 \\ \sigma_2 \\ \sigma_1 \\ \rho \\ \gamma_v \\ \gamma_u \end{pmatrix},$$

where we defined $\gamma_v = -v + \gamma\mu s_v^{-1}$ and $\gamma_u = -u + \gamma\mu s_u^{-1}$.

After elimination of the unknowns

$$\begin{array}{ll} d_{s_u} = \sigma_2 + d_\lambda\,, & d_{s_v} = \sigma_3 - d_\lambda\,, \\ d_u = \gamma_u - s_u^{-1}u d_{s_u}\,, & d_v = \gamma_v - s_v^{-1}v d_{s_v}\,; \end{array}$$

$(23)_\gamma$ reduces to

$$\begin{pmatrix} -\frac{\alpha}{m}I & A^T \\ A & D \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} = \begin{pmatrix} \sigma_1 \\ \tilde{\rho} \end{pmatrix} \tag{24}$$

where $D = diag(s_u^{-1}u + s_v^{-1}v)$ and $\tilde{\rho} = \rho - s_u^{-1}u\sigma_2 + \gamma_u + s_v^{-1}v\sigma_3 - \gamma_v$. Observe that all the change along iterations is in D, together with the righthand side, of course. Note also that none of the eliminations have produced off-diagonal fill-in in the remaining system. However, proceeding further will definitely introduce such fill-in.

To solve (24), we solve first a reduced system for d_λ as follows. The second equation in (24) gives $d_\lambda = D^{-1}(\tilde{\rho} - Ad_x)$. Plug this value in the first equation to obtain

$$(\frac{\alpha}{m}I + A^T D^{-1} A)d_\lambda = \sigma_1 + A^T D^{-1}\tilde{\rho}\,,$$

a normal-equations like system for d_λ. The main advantage of this approach is that the reduced matrix is positive definite with constant sparsity pattern along iterations. This feature allows us to choose in the first iteration, once and for ever, a permutation matrix P such that the pivot order is preserved: $P(\frac{\alpha}{m}I + A^T D^{-1} A)P^T$, and whose Cholesky factor is more sparse than the reduced matrix $\frac{\alpha}{m}I + A^T D^{-1} A$.

5.1 Starting point

The choice of the initial values x^0, u^0, v^0, λ^0, s_u^0, s_v^0 and μ_0 is done as follows.

- For x^0, it is natural to take z, the noisy image.
- The values of u^0 and v^0 are chosen to have primal feasibility:

$$u^0 = \max\{Ax^0, 0\} \quad \text{and} \quad v^0 = \max\{-Ax^0, 0\}\,.$$

- For λ^0 we distinguish three cases
  - If $A(i,.)x > 0$ take $\lambda_i^0 = 1$. Let I_1 be the corresponding index set.
  - If $A(i,.)x < 0$ take $\lambda_i^0 = -1$. Let I_2 be the corresponding index set.
  - For the rest of indices, $i \in (I_1 \cup I_2)^c$, solve

$$(A(i,.))^T_{i\in(I_1\cup I_2)^c}\lambda_{(I_1\cup I_2)^c} = -(p + \frac{\alpha}{m}x^0 + \sum_{i\in I_1\cup I_2}\lambda_i^0 A(i,.)^T)\,.$$

- Having λ^0, project it onto $[-1,1]^r$ and take

$$s_{u^0} = \max\{\lambda^0 + e, 0\} \quad \text{and} \quad s_{v^0} = \max\{e - \lambda^0, 0\}\,.$$

- Finally, we choose $\mu_0 = ((u^0)^T s_{u^0} + (v^0)^T s_{v^0})/(2r)$.

In order to ensure positivity in u, v, s_u and s_v, we replace by 1 the respective components smaller than 1, *pushing* variables towards positive values, sufficiently bounded away from zero.

5.2 Stopping rule

Although the stopping criterion in STEP 4 of (PCIP) is $\mu_k \leq \mu_\infty$, interior point algorithms rather verify when optimality conditions (22) are satisfied, up to some tolerance. This stopping test is stronger than the one in STEP 4, see Table 1.

For infeasible predictor corrector methods, approximate satisfaction of optimality conditions means (up to a relative tolerance) to have primal and dual feasibility and zero duality gap in (5), see [8]. Then, for given tolerances TOL_1, $\text{TOL}_2 > 0$ we check

$$\begin{cases} \text{Primal feasibility} & \|Ax - u + v\|_\infty \leq \text{TOL}_1 \\ \text{Dual feasibility} & \dfrac{\|(-\frac{\alpha}{m}x + A^T\lambda - p, -\lambda + s_u - e, \lambda + s_v - e)\|_\infty}{\|(p,1,1)\|_\infty + 1} \leq \text{TOL}_1 \\ \text{No duality gap} & \dfrac{u^T s_u + v^T s_v}{|F(x,u,v)| + 1} \leq \text{TOL}_2\,, \end{cases} \tag{25}$$

where $F(x,u,v)$ stands for the objective in (5).

6 Bundle method

We already mentioned that (4) is a nonsmooth problem. More precisely, let $f : \mathbb{R}^m \to \mathbb{R}$ be the objective function in (4):

$$f(x) = \frac{\alpha}{2m}x^T x + \frac{\alpha}{m}z^T x + \|Ax\|_1\,,$$

where A is the matrix from (3). Then (4) can be solved using a nondifferentiable optimization method, of dimension N^2 if the image has $N \times N$ pixels. We outline here a Reversal quasi-Newton Bundle algorithm which is explained in full details in [5, § 5].

Assume that an *oracle* computes $f(y)$ and a subgradient $g(y)$ in the subdifferential $\partial f(y)$, for any given y. After k iterations, the *cutting-plane* model of f is built:

$$\check{f}_k(y) := \max\{f^i + (y - y^i)^T g^i \ : \ i = 1, 2, \ldots, k\}\,, \tag{26}$$

where $(y^i,\ f^i = f(y^i),\ g^i = g(y^i) \in \partial f(y^i), i = 1, \ldots, k)$ form the bundle.

The bundle mechanism consists in generating two sequences: first, a sequence of sampling points $\{y^i\}^i$, to construct the model (26). From these points, a subsequence of *serious* points $\{x^j\}$ is selected. They aim at guaranteeing a sufficient decrease in the original function f: y^k is declared a serious step ($x^{j+1} = y^k$) when

$$f(y^k) \leq f(x^j) - \alpha_1 \delta_k\,, \tag{27}$$

where $\alpha_1 \in]0, 1[$ is an Armijo-like tolerance and δ_k is the *nominal decrease* predicted by the model.

When a sampling point does not satisfy (27), the model $\check{f}_k$ is considered not accurate enough and a *null step* is declared: there is no new serious point and a new sampling point is generated.

It is possible to merge this bundling technique with a curve-search. This is subalgorithm (CS) in [5]. Given x^j and a positive *poor-man's quasi-Newton* parameter $m_j > 0$, the idea is to generate sampling points by solving

$$\min_{y \in \mathbb{R}^m} \check{f}_k(y) + \frac{1}{2}\frac{m_j}{t}(y - x^j)^T(y - x^j)\,, \tag{28}$$

for trial stepsizes $t > 0$. Associated to a solution $y(t)$ of (28) there is the quantity

$$\delta(t) = f(x^j) - \check{f}_k(y(t)) - \frac{1}{2}\frac{m_j}{t}(y(t) - x^j)^T(y(t) - x^j)\,.$$

The possible exits of (CS) are: serious step, null step and an emergency stop, when the cutting-plane function has been minimized. A *cutting-plane step* copes with the case when (27) holds, but t starts going to ∞, because $y(t)$ minimizes $\check{f}_k$ and becomes constant. For the sake of simplicity, we omit entering into more technical details here.

One word about the stopping test. Together with each sampling point $y(t)$, we generate a particular subgradient

$$\check{G}(t) = \frac{m_j}{t}(x^j - y(t)) \in \partial \check{f}_k(y(t))$$

satisfying the relation $\check{G}(t) \in \partial_{\check{e}(t)} f(x^j)$, where $\check{e}(t) = f(x^j) - \check{f}_k(y(t)) + \check{G}(t)^T(y(t) - x^j)$ is a "linearization" error and $\partial_\varepsilon f$ is the ε-subdifferential from Convex Analysis. When both $\|\check{G}(t)\|$ and $\check{e}(t)$ are small, the inclusion $0 \in \partial f(x^j)$ is approximately satisfied. From this observation, the algorithm stops when

$$\|\check{G}(t)\| \leq \text{TOL}_3 \quad \text{and} \quad \check{e}(t) \leq \text{TOL}_4\,, \tag{29}$$

where TOL3 and TOL4 are given tolerances.

Altogether, the algorithmic scheme is as follows.

Reversal quasi-Newton Bundle Algorithm (RQB)

STEP 0. The initial x^1 and m_1 are given, as well as stopping and curve-search tolerances. Define the initial cutting-plane model $\check{f}_1$ according to (26) and set $k = j = 1$.

STEP 1. Set $t = 1$ and call the curve-search subalgorithm (CS). If (29) does not hold, it exits with a new $t_k > 0$, a sampling point $y^{k+1} = y(t_k)$, the nominal decrease $\delta_k = \delta(t_k)$ and a message indicating

–either null-step, when (27) does not hold

–or cutting-plane-step, when (27) holds, but $y(t)$ becomes constant,

–or serious-step.

STEP 2. In case of null-step, go to Step 3.

Otherwise update $x^{j+1} = y^{k+1}$.

In case of serious-step, set $\zeta := x^{j+1} - x^j$. Compute $v = g(x^{j+1}) - g(x^j)$ and update m_{j+1}, using the formula (see § 4.2 in [5])

$$\frac{1}{m_{j+1}} = \frac{1}{m_j} + \frac{v^T\zeta}{||v||^2}\,.$$

Replace j by $j+1$.

STEP 3. Update the cutting-plane model $\check{f}_{k+1}$ of (26). Replace k by $k+1$ and loop to Step 1. □

The code (RQB) is part of Inria's MODULOPT library. It is distributed for free for academic applications.

7 Numerical results

We have implemented (PCIP) in MATLAB and (RQB) in FORTRAN, on a SPARC 10, with 80MB of memory and running under SunOS 4.1.3.

For our comparisons we generated a set of five problems, starting from the blocky image of Fig. 1, going to the smooth, bell-shaped example shown in Fig. 5. Figs. 2-4 correspond to in-between smoothings of the sharp faces and edges in Fig. 1). The regularization by bounded-variations seminorms is expected to be especially effective for blocky images.

Each example starts from an original image x^* defined over a square with 80 pixels on each side. Therefore $N = 80$, and the dimensions treated by (PCIP) and (RQB) are 69600 and 6400, respectively. To obtain an initial noisy image z, we perturb x^* with random uniformly distributed numbers. Each figure contains 4 graphics. On the top, x^* is on the left and z is shown on the right. Images obtained with (PCIP) and (RQB) appear in the bottom, left and righthand sides, respectively, with their CPU times.

Example	PCIP: $\mu_{\bar{k}}$	PCIP: Error	PCIP: CPU	RQB: Error	RQB: CPU
#1	7E-14	0.10	3m25s	0.11	3m47s
#2	2E-11	0.12	3m32s	0.12	3m26s
#3	3E-11	0.15	3m20s	0.15	3m44s
#4	2E-11	0.15	2m55s	1.45	3m14s
#5	2E-13	0.41	3m45s	20.4	3m14s

Table 1: Numerical results in figures

Table 1 summarizes our numerical results. For each example we list (PCIP) and (RQP)'s performances, measured in CPU times and in terms of errors. By (relative) error we mean the following:

$$\frac{\|x^{ap} - x^*\|_\infty}{\max(1, \|x^*\|_\infty)},$$

where x^{ap} is the numerical solution. Beware that x^* is **not** the solution of the discretized problem, but rather of the infinite dimensional one. However, since the original image is available in our examples, we use it to estimate the error.

We also show the current values of $\mu_{\bar{k}}$ in (PCIP), such that the stopping test (25) holds. For (PCIP) we use the tolerances $\text{TOL}_1 = 10^{-6}$ and $\text{TOL}_2 = 10^{-8}$, while in (RQB) we chose $\text{TOL}_3 = 10^{-5}$ and $\text{TOL}_4 = 10^{-8}$.

Observe that the smoother is the image, the worse is the relative error. This is natural, since the seminorm regularization approach is more suitable for blocky images. When comparing CPU times, we see that both algorithms obtain similar numerical solutions in comparable times for the first three examples. However, (PCIP) behaves better than (RQB) for recovering the smoother images.

8 Concluding remarks

We presented an infeasible predictor-corrector interior point method for solving monotone linear complementarity problems. The method has good convergence properties, such as polynomial complexity of order $O(n\sqrt{n}L)$ and quadratic asymptotic rate. Moreover, it can be effectively implemented to have good numerical performances, provided there is some underlying sparsity, like in the image recovery case.

References

[1] F. Bonnans and F. A. Potra. Infeasible path following algorithms for linear complementarity problems. *Mathematics of Operations Research* 22-2, 378–407, 1997.

[2] K. Kunisch, E. Casas and C. Pola. Regularization by functions of bounded variation and applications to image enhancement. Technical Report 1996.5, Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain, 1996.

[3] C.C. Gonzaga and J.F. Bonnans. Convergence of interior point algorithm for the monotone linear complementarity problem. *Mathematics of Operations Research*, 21(1996), pp. 1-25.

[4] F. Bonnans, C. Pola and R. Rebai. Perturbed path following interior point algorithms. Technical Report 2745, INRIA, France, 1995.

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

[6] S. Mizuno. A new polynomial time method for a linear complementarity problem. *Mathematical Programming*, 56:31–43, 1992.

[7] J. Stoer S. Mizuno, F. Jarre. A unified approach to infeasible-interior-point algorithms via geometrical linear complementarity problems. Technical Report 213, Math. Inst. Univ. Wurzburg, Germany, 1994.

[8] R.J. Vanderbei. LOQO: An interior point code for quadratic programming. Technical Report SOR 94-15, Princeton University, 1995.

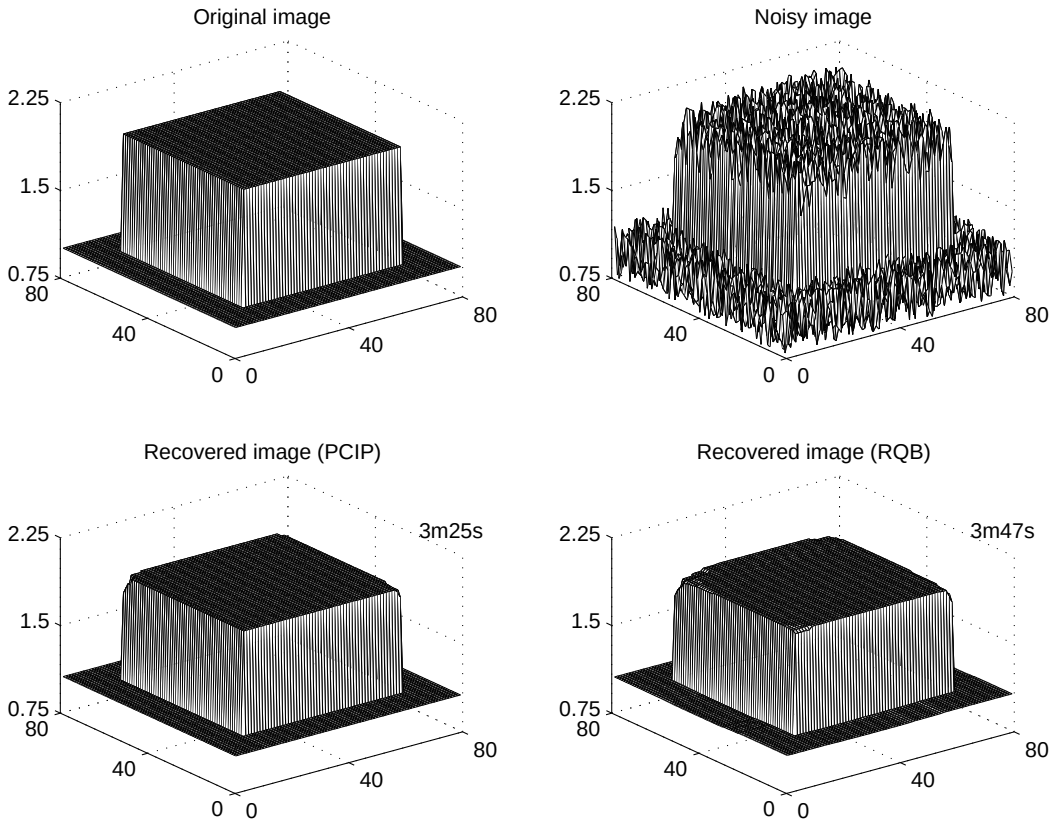

Figure 1: CPU times - Example 1

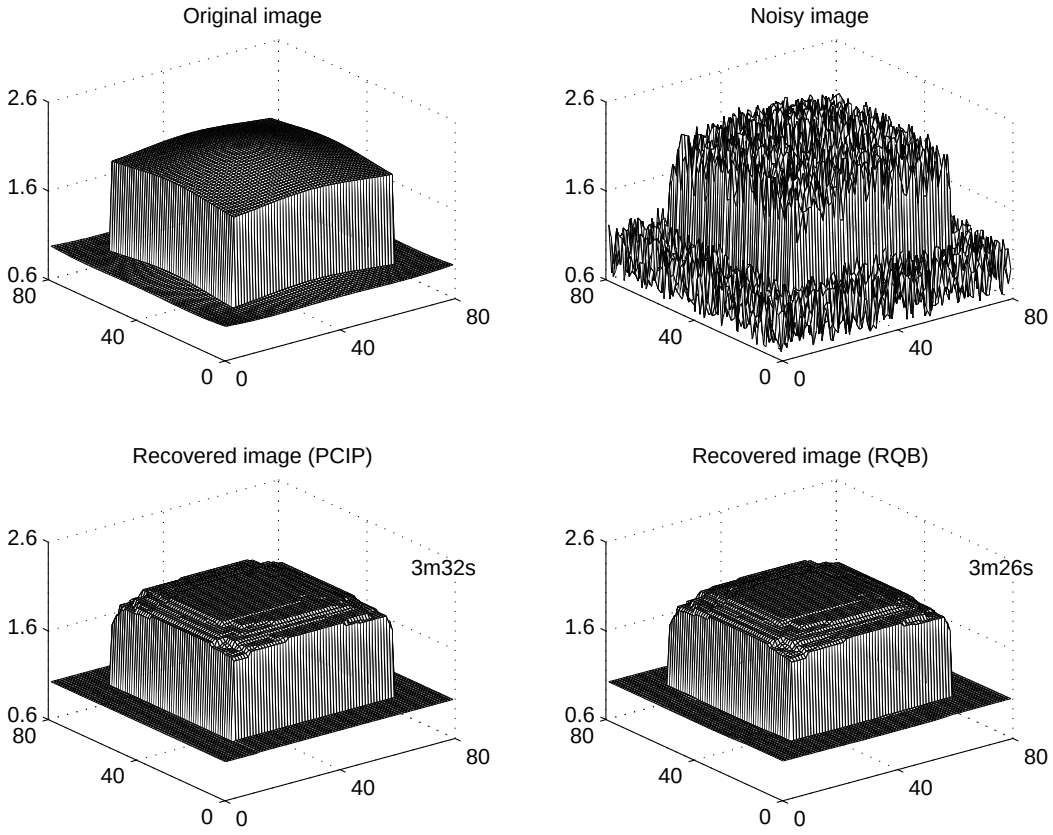

Figure 2: CPU times - Example 2

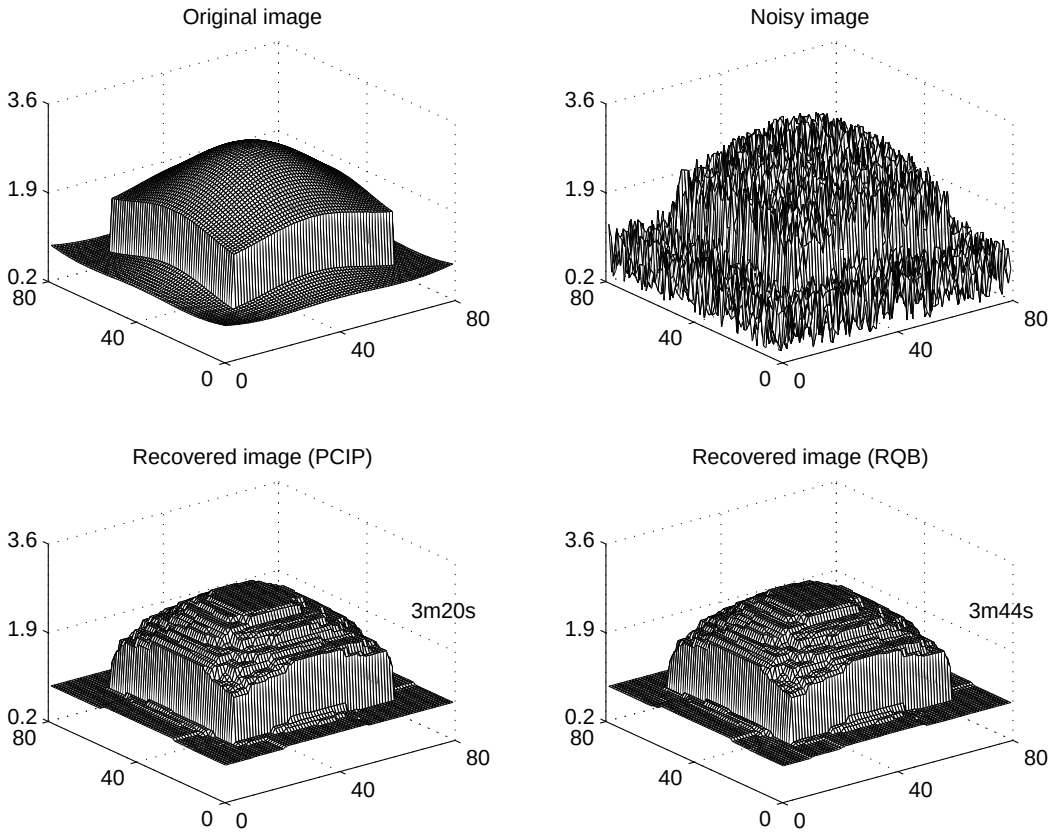

Figure 3: CPU times - Example 3

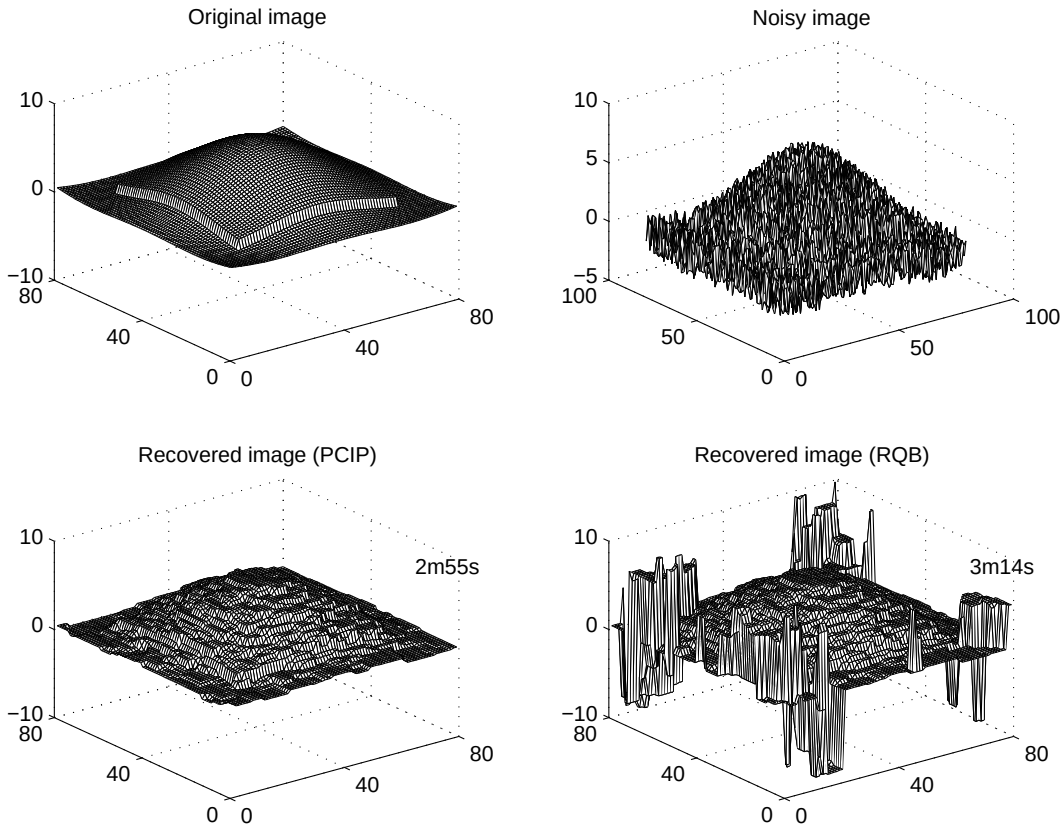

Figure 4: CPU times - Example 4

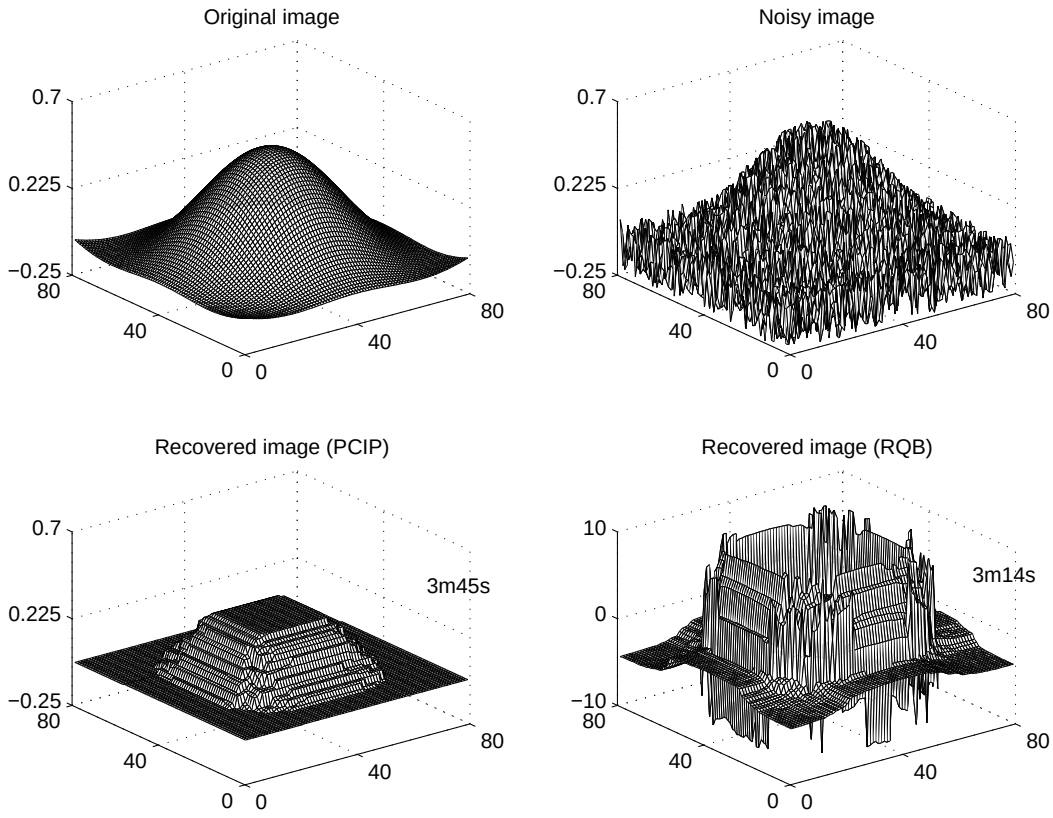

Figure 5: CPU times - Example 5

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l' Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399