

strained problems where the objective function is the sum of a smooth term and a convex separable term (the latter might be an indicator function that enforces bound constraints), or allow for a non-separable structure but require full gradient evaluations to choose the working set, or have convergence results in expectation. In particular, active-set identification results are given in [48] for variants of the sequential minimal optimization algorithm applied to the Support Vector Machine problem, where the authors consider a random selection of the working-set, which therefore does not require first-order information, but leads to convergence results in expectation.

- (ii) Besides stating finite active-set identification results in a general non-convex setting, complexity results are also given under convexity of the objective function and a quadratic growth condition (satisfied by any strongly convex function), allowing us to bound the maximum number of iterations needed to identify the active set.

Let us also remark that here we consider a simple Armijo line search for computing the stepsize along any search direction, thus not requiring exact minimizations, or the knowledge of the Lipschitz constant of the gradient, or other additional information. This makes our analysis of particular interest for realistic application to large-scale optimization problems.

2. Preliminaries and Notation. Let us first introduce part of the notation used in the paper. Given a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, we indicate the gradient of f by ∇f and we denote by $\nabla_i f$ its i th component (i.e., the i th partial derivative of f). For a vector $x \in \mathbb{R}^n$, we denote by x_i the i th component of x , we indicate by $\|x\|$ the Euclidean norm of x and we indicate by $\|x\|_\infty$ the sup-norm of x . We also denote by $e \in \mathbb{R}^n$ the vector made of all ones, and by $e_i \in \mathbb{R}^n$ the vector that has the i th component equal to 1 and all other components equal to 0. Given a scalar a , we indicate with $\lfloor a \rfloor$ the largest integer less than or equal to a .

Our analysis is concerned with the following problem:

$$(2.1) \quad \begin{aligned} & \min f(x) \\ & e^T x = b \\ & l_i \leq x_i \leq u_i, \quad i = 1, \dots, n, \end{aligned}$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a function with Lipschitz continuous gradient, $n \geq 2$, $b \in \mathbb{R}$ and, for all $i = 1, \dots, n$, we have $l_i < u_i$, $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$. The feasible set of problem (2.1) is denoted by \mathcal{F} .

Note that we may consider, instead of $e^T x = b$, any constraint of the form $a^T x = b$, with $a_i \neq 0$, $i = 1, \dots, n$. In such a case, problem (2.1) can be obtained by applying the variable transformation $x_i \leftarrow a_i x_i$ and setting the lower and the upper bound accordingly. (Examples of relevant applications where problem (2.1) arises can be found, e.g., in [12] and the references therein.)

The Lipschitz constant of ∇f over \mathbb{R}^n is denoted by L , that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

It is possible to show [1] that there exist local Lipschitz constants

$$(2.2) \quad L_{i,j} \leq 2L, \quad i, j = 1, \dots, n,$$

such that, for any $x \in \mathbb{R}^n$,

$$|\nabla f(x + s(e_i - e_j))^T(e_i - e_j) - \nabla f(x + t(e_i - e_j))^T(e_i - e_j)| \leq L_{i,j}|s - t|, \quad \forall s, t \in \mathbb{R}.$$

87 Equivalently, defining $\phi_{i,j,x}(\alpha) = f(x + \alpha(e_i - e_j))$ and denoting its derivative by
 88 $\dot{\phi}_{i,j,x}$, we have that

$$89 \quad (2.3) \quad |\dot{\phi}_{i,j,x}(s) - \dot{\phi}_{i,j,x}(t)| \leq L_{i,j}|s - t|, \quad \forall s, t \in \mathbb{R},$$

90 that is, each derivative $\dot{\phi}_{i,j,x}$ is Lipschitz continuous over \mathbb{R} with constant $L_{i,j}$.

91 Without loss of generality, we assume that all $L_{i,j} > 0$, $i \neq j$ (if some of them
 92 are equal to zero, they can be replaced by positive overestimates) and that $L_{i,i} = 0$,
 93 $i = 1, \dots, n$. We also define the following constants:

$$94 \quad (2.4) \quad L^{\max} = \max_{i,j=1,\dots,n} L_{i,j},$$

$$95 \quad (2.5) \quad L_j = \sum_{i=1}^n L_{i,j}, \quad j = 1, \dots, n,$$

$$96 \quad (2.6) \quad \hat{L}^{\max} = \max_{j=1,\dots,n} L_j.$$

98 A characterization of stationary points for problem (2.1) follows from KKT con-
 99 ditions. In particular, a point $x^* \in \mathcal{F}$ is stationary for problem (2.1) if and only if
 100 there exists $\lambda^* \in \mathbb{R}$ such that, for all $i = 1, \dots, n$,

$$101 \quad (2.7) \quad \nabla_i f(x^*) \begin{cases} \geq \lambda^*, & \text{if } x_i^* = l_i, \\ = \lambda^*, & \text{if } x_i^* \in (l_i, u_i), \\ \leq \lambda^*, & \text{if } x_i^* = u_i. \end{cases}$$

102 Moreover, a variable $x_i^* \in \{l_i, u_i\}$ is said to satisfy the strict complementarity if
 103 $\nabla_i f(x^*) \neq \lambda^*$. We also say that x^* is non-degenerate if all variables x_i^* such that
 104 $x_i^* \in \{l_i, u_i\}$ satisfy the strict complementarity.

105 In the following, we will make use of a simple operator between vectors in \mathbb{R}^n , ob-
 106 tained from the usual dot product by discarding a certain component. More precisely,
 107 for any $j \in \{1, \dots, n\}$ we define the following positive semidefinite inner product:

$$108 \quad \langle x, y \rangle_j = \sum_{i \neq j} x_i y_i, \quad \forall x, y \in \mathbb{R}^n.$$

109 We also define the following seminorm, induced by the above inner product:

$$110 \quad \|x\|_{\langle j \rangle} = \sqrt{\langle x, x \rangle_j}, \quad \forall x \in \mathbb{R}^n.$$

111 Note that, by Cauchy-Bunyakovsky-Schwarz inequality, we have

$$112 \quad (2.8) \quad \langle x, y \rangle_j \leq \|x\|_{\langle j \rangle} \|y\|_{\langle j \rangle}, \quad \forall x, y \in \mathbb{R}^n.$$

113 In particular, (2.8) implies that

$$114(2.9) \quad |x_i| \leq \|x\|_{\langle j \rangle}, \quad i \neq j, \quad \forall x \in \mathbb{R}^n,$$

$$115(2.10) \quad \sum_{i \neq j} |x_i| \leq \sqrt{n-1} \|x\|_{\langle j \rangle}, \quad \forall x \in \mathbb{R}^n.$$

116

117 Moreover, it is straightforward to verify that

$$118 \quad (2.11) \quad \|x\|_{\langle j \rangle} \leq \|x\|, \quad \forall x \in \mathbb{R}^n.$$

119 **3. Review of the algorithm.** Let us briefly review the algorithm proposed
 120 in [12], named Almost Cyclic 2-Coordinate Descent (AC2CD) method, to solve prob-
 121 lem (2.1). The main feature of AC2CD is an *almost cyclic* rule to choose the working
 122 set. This rule iteratively selects two variables: one is picked in a cyclic fashion, while
 123 the other one is chosen by considering the distance from the bounds in some points
 124 produced by the algorithm and remains in the working set until all the other variables
 125 have been picked. Note the difference from the so-called *essentially cyclic* rule, where
 126 all blocks of variables must be selected at least once within a certain number of steps.

127 More precisely, at the beginning of each outer iteration k of AC2CD we have a
 128 feasible point x^k and we select a variable index $j(k)$ such that $x_{j(k)}^k$ is “sufficiently
 129 far” from its nearest bound. Then, we set the point $z^{k,1} = x^k$ and start a cycle of
 130 inner iterations, which are denoted by $(k, 1), \dots, (k, n)$. In each inner iteration (k, i) ,
 131 we choose a working set of two variables: one of them is selected in a cyclic fashion,
 132 while the other one remains the $j(k)$ th variable. So, we produce a feasible point $z^{k,i+1}$
 133 from $z^{k,i}$ by moving only the two variables in the working set. At the end of the last
 134 inner iteration we finally set $x^{k+1} = z^{k,n+1}$ and start a new outer iteration $k + 1$.

135 Let us remark that our algorithm does not use first-order information to choose
 136 the working set. Moreover, as to be described later, only two partial derivatives are
 137 required to move each pair of variables. We can hence achieve high computational
 138 efficiency if partial derivative evaluation for the objective function is much cheaper
 139 than full gradient evaluation. For instance, this is the case when f is the sum of
 140 univariate functions (such as in the problems considered in [38] for large-scale network
 141 optimization). Other interesting examples, including the Support Vector Machine
 142 problem and the Chebyshev center problems, are those where the objective function
 143 is quadratic of the form $f(x) = x^T Q^T Q x - q^T x$, with Q being a given $m \times n$ matrix
 144 and q being a given vector. In this case, a partial derivative of $f(x)$ can be computed
 145 with a cost $\mathcal{O}(m)$, while computing the whole gradient has a cost $\mathcal{O}(mn)$ (see [12] for
 146 details).

147 Now, let us explain in more detail how the index $j(k)$ is chosen at the beginning
 148 of an outer iteration k and how the two variables in the working set are moved in the
 149 inner iterations $(k, 1), \dots, (k, n)$.

150 For what concerns the choice of $j(k)$, for any $x \in \mathcal{F}$ let us first define

$$151 \quad (3.1) \quad D_h(x) = \min\{x_h - l_h, u_h - x_h\}, \quad h = 1, \dots, n.$$

152 Namely, $D_h(x)$ returns the distance of x_h from its nearest bound. Moreover, for any
 153 point x^k produced by the algorithm, we define D^k as the maximum distance between
 154 each component of x^k and its nearest bound, that is,

$$155 \quad (3.2) \quad D^k = \max_{h=1, \dots, n} D_h(x^k).$$

156 Then, $j(k)$ can be chosen as any index satisfying

$$157 \quad (3.3) \quad D_{j(k)}(x^k) \geq \tau D^k,$$

158 where $\tau \in (0, 1]$ is a fixed parameter. In other words, the distance between $x_{j(k)}^k$ and
 159 its nearest bound must be sufficiently large compared to D^k .

160 For what concerns the variable update, let us denote by p_i^k the variable index
 161 that is selected in a cyclic manner at an inner iteration (k, i) (note that the variables
 162 can be taken in any order). So, $z_{p_i^k}^{k,i}$ and $z_{j(k)}^{k,i}$ are the two variables that can be moved

163 from $z^{k,i}$. To do this, we use the following search direction (which has at most two
 164 non-zero components and maintains feasibility for the equality constraint):

$$165 \quad (3.4) \quad d^{k,i} = g^{k,i}(e_{p_i^k} - e_{j(k)}), \quad \text{where} \quad g^{k,i} = \nabla_{j(k)} f(z^{k,i}) - \nabla_{p_i^k} f(z^{k,i}),$$

166 and we set

$$167 \quad z^{k,i+1} = z^{k,i} + \alpha^{k,i} d^{k,i},$$

168 where $\alpha^{k,i}$ is a suitably computed feasible stepsize. Note that

$$169 \quad (3.5) \quad \nabla f(z^{k,i})^T d^{k,i} = -(g^{k,i})^2,$$

170 and then, every non-zero $d^{k,i}$ is a descent direction. The scheme of AC2CD is reported
 171 in Algorithm 3.1.

Algorithm 3.1 Almost Cyclic 2-Coordinate Descent (AC2CD) method

```

0  Given  $x^0 \in \mathcal{F}$  and  $\tau \in (0, 1]$ 
1  For  $k = 0, 1, \dots$ 
2    Choose a variable index  $j(k) \in \{1, \dots, n\}$  that satisfies (3.3)
3    Choose a permutation  $\{p_1^k, \dots, p_n^k\}$  of  $\{1, \dots, n\}$ 
4    Set  $z^{k,1} = x^k$ 
5    For  $i = 1, \dots, n$ 
6      Let  $g^{k,i} = \nabla_{j(k)} f(z^{k,i}) - \nabla_{p_i^k} f(z^{k,i})$ 
7      Compute the search direction  $d^{k,i} = g^{k,i}(e_{p_i^k} - e_{j(k)})$ 
8      Compute a feasible stepsize  $\alpha^{k,i}$  and set  $z^{k,i+1} = z^{k,i} + \alpha^{k,i} d^{k,i}$ 
9    End for
10   Set  $x^{k+1} = z^{k,n+1}$ 
11 End for

```

172 **3.1. Computation of the stepsize.** Under a technical assumption (see As-
 173 sumption 1 in the next section), global convergence of AC2CD to stationary points
 174 was established in [12] for different choices of the stepsize $\alpha^{k,i}$ (to be used at line 8
 175 of Algorithm 3.1), including the Armijo stepsize, overestimates of the local Lipschitz
 176 constants of ∇f and the exact stepsize for strictly convex objective functions¹.

177 Here we focus on the case where, at every inner iteration (k, i) , the stepsize $\alpha^{k,i}$
 178 is computed by the Armijo line search, which is a backtracking procedure that computes
 179 a stepsize in a finite number of iterations. The scheme of the Armijo line search used
 180 in AC2CD is reported in Algorithm 3.2.

Algorithm 3.2 Armijo line search (to compute $\alpha^{k,i}$ at step 8 of AC2CD)

```

0  Given the search direction  $d^{k,i}$  and two parameters  $\gamma \in (0, 1)$ ,  $\delta \in (0, 1)$ 
1  Choose a feasible stepsize  $\Delta^{k,i} \geq 0$  and set  $\alpha = \Delta^{k,i}$ 
2  While  $f(z^{k,i} + \alpha d^{k,i}) > f(z^{k,i}) + \gamma \alpha \nabla f(z^{k,i})^T d^{k,i}$ 
3    Set  $\alpha = \delta \alpha$ 
4  End while
5  Return  $\alpha^{k,i} = \alpha$ 

```

¹For general conditions on the stepsize, see SC (Stepsize Condition) 1 in [12]. A typo is present in point (i) of SC 1 in [12]: $f(z^{k,i+i})$ should be replaced by $f(z^{k,i+1})$.

181 We see that the considered Armijo line search is very simple and does not require
 182 exact minimizations or additional information (such as the knowledge of the Lipschitz
 183 constant of ∇f). For this reason, it can be an effective choice for non-convex large-
 184 scale problems and when no closed form is known for the stepsize.

185 To obtain global convergence of AC2CD to stationary points, an appropriate
 186 choice of the initial stepsize $\Delta^{k,i}$ at line 1 of Algorithm 3.2 is needed. In [12] it was
 187 shown that, at every inner iteration (k, i) , a possible choice is the following:

$$188 \quad (3.6) \quad \Delta^{k,i} = \min\{\bar{\alpha}^{k,i}, A^{k,i}\},$$

189 where

- 190 • $\bar{\alpha}^{k,i}$ is the largest feasible stepsize along the direction $d^{k,i}$, that is,

$$191 \quad (3.7) \quad \bar{\alpha}^{k,i} = \begin{cases} \frac{1}{g^{k,i}} \min\{u_{p_i^k} - z_{p_i^k}^{k,i}, z_{j(k)}^{k,i} - l_{j(k)}\}, & \text{if } g^{k,i} > 0, \\ \frac{1}{|g^{k,i}|} \min\{z_{p_i^k}^{k,i} - l_{p_i^k}, u_{j(k)} - z_{j(k)}^{k,i}\}, & \text{if } g^{k,i} < 0, \\ 0, & \text{if } g^{k,i} = 0; \end{cases}$$

- 192 • $A^{k,i}$ must be chosen between two finite positive constants, that is,

$$193 \quad (3.8) \quad 0 < A_l \leq A^{k,i} \leq A_u < \infty,$$

194 with A_l and A_u being two fixed parameters.

195 We observe that, in (3.7), we set $\bar{\alpha}^{k,i} = 0$ when $g^{k,i} = 0$, i.e., when $d^{k,i} = 0$ (see (3.4)).
 196 Therefore, $\bar{\alpha}^{k,i}$ is not actually the largest feasible stepsize along $d^{k,i}$ when $d^{k,i} = 0$.
 197 This choice in the definition of $\bar{\alpha}^{k,i}$ simplifies the analysis and entails no loss of
 198 generality, since it stills guarantees that $z^{k,i+1} = z^{k,i}$ when $d^{k,i} = 0$. In particular,
 199 note that

$$200 \quad (3.9) \quad d^{k,i} = 0 \stackrel{(3.4)}{\Leftrightarrow} g^{k,i} = 0 \stackrel{(3.7)}{\Rightarrow} \bar{\alpha}^{k,i} = 0 \Leftrightarrow z^{k,i+1} = z^{k,i}.$$

201 To obtain the last relation in (3.9), we can use (3.5), (3.6) and (3.8), leading to

$$202 \quad \bar{\alpha}^{k,i} > 0 \Leftrightarrow \Delta^{k,i} > 0 \wedge \nabla f(z^{k,i})^T d^{k,i} < 0.$$

203 So, if $\bar{\alpha}^{k,i} > 0$, the Armijo line search returns a stepsize $\alpha^{k,i} > 0$, implying that
 204 $z^{k,i+1} \neq z^{k,i}$. Vice versa, if $\bar{\alpha}^{k,i} = 0$, the Armijo line search returns $\alpha^{k,i} = 0$,
 205 implying that $z^{k,i+1} = z^{k,i}$. Namely, the last relation in (3.9) holds.

206 **4. Basic assumptions.** Let X^* be the set of all stationary points for prob-
 207 lem (2.1) and also define the level set

$$208 \quad \mathcal{L}^0 = \{x \in \mathcal{F}: f(x) \leq f(x^0)\},$$

209 where \mathcal{F} is the feasible set of problem (2.1) and x^0 is the starting point used in AC2CD.
 210 We assume that \mathcal{L}^0 is non-empty and compact (implying that both the feasible set \mathcal{F}
 211 and the set of stationary points X^* are non-empty as well).

212 According to the results stated in [12], we also need the following assumption on
 213 the level set \mathcal{L}^0 to ensure global convergence of AC2CD (in the sense that every limit
 214 point of the sequence $\{x^k\}$ produced by the algorithm is stationary):

215 ASSUMPTION 1. $\forall x \in \mathcal{L}^0, \exists i \in \{1, \dots, n\}: x_i \in (l_i, u_i)$.

216 Namely, we require that every point of \mathcal{L}^0 has at least one component strictly between
 217 the lower and the upper bound. Note that Assumption 1 is automatically satisfied
 218 when \mathcal{F} is the unit simplex (i.e., when in problem (2.1) we have $b = 1$, $l_i = 0$, $u_i = +\infty$,
 219 $i = 1, \dots, n$). Moreover, in [33] it is shown that Assumption 1 is also satisfied for the
 220 Support Vector Machine training problem if $f(x^0) < 0$ and the smallest eigenvalue of
 221 the Hessian matrix of $f(x)$ is sufficiently large. (Assumption 1 is satisfied also when
 222 at least one variable has are no finite bounds, provided \mathcal{F} is not a singleton.)

223 Essentially, Assumption 1 is needed to prevent AC2CD from converging to a
 224 point x^* with all components at the lower or the upper bound. To be more spe-
 225 cific, the convergence analysis of AC2CD (see [12]) relies on the fact that eventually
 226 $l_{j(k)} < x_{j(k)}^k < u_{j(k)}$ and that $\nabla_{j(k)} f(x^k)$ converges (over suitable subsequences) to
 227 the KKT multiplier λ^* appearing in (2.7). Also the analysis of the active-set identi-
 228 fication reported later uses the same properties (see Proposition 6.2 and the proof of
 229 Theorem 6.4 below). Without Assumption 1, all these results do not hold, since $\{x^k\}$
 230 may have limit points with all components at the lower or the upper bound.

231 We also observe that, for every outer iteration $k \geq 0$, Assumption 1 ensures that
 232 $x^{k+1} \neq x^k$ if and only if x^k is non-stationary. To see this, under Assumption 1 observe
 233 that $l_{j(k)} < x_{j(k)}^k < u_{j(k)}$ for all $k \geq 0$ (since $j(k)$ must satisfy (3.3) with $D^k > 0$).
 234 Then, from the KKT conditions (2.7), there exists a feasible descent direction in the
 235 inner iterations $(k, 1), \dots, (k, n)$ if and only if x^k is non-stationary.

236 On the contrary, without Assumption 1, the algorithm may end up in a non-
 237 stationary point x^k with all components at the lower or the upper bound. In such a
 238 case, even if every choice of $j(k) = 1, \dots, n$ satisfies (3.3) (since $D^k = D_h(x^k) = 0$,
 239 $h = 1, \dots, n$), for certain choices of $j(k)$ there may not exist a feasible descent direction
 240 in any inner iteration $(k, 1), \dots, (k, n)$. Namely, AC2CD may get stuck in a non-
 241 stationary point x^k . This issue can be overcome by introducing an anticycling rule to
 242 select $j(k)$ when such a point x^k is produced. Doing so, we may relax Assumption 1
 243 by requiring only the stationary points in \mathcal{L}^0 not to have all components at the lower
 244 or the upper bound, but in our analysis we use Assumption 1 for simplicity.

245 Overcoming the limitation deriving from Assumption 1 by properly modifying
 246 the algorithm might be a challenging subject for future research.

247 In the rest of the paper, we will consider all the above assumptions always satisfied,
 248 even if not explicitly invoked. Namely, we will consider \mathcal{L}^0 non-empty and compact
 249 and we will consider Assumption 1 satisfied.

250 **5. Technical results.** In this section, we fix a few concepts and give some
 251 technical results. First note that, for every inner iteration (k, i) of AC2CD,

$$252 \quad (5.1) \quad p_i^k \neq j(k) \quad \Rightarrow \quad z_{p_i^k}^{k,i} = x_{p_i^k}^k \quad \text{and} \quad x_{p_i^k}^{k+1} = z_{p_i^k}^{k,i+1},$$

253 since each coordinate, except the $j(k)$ th one, is moved (at most) once in a cycle of
 254 inner iterations.

255 Furthermore, there is a relation between the Armijo stepsize and the local Lips-
 256 chitz constants of ∇f : at any inner iteration (k, i) , every stepsize $\alpha \leq 2(1-\gamma)/L_{p_i^k, j(k)}$
 257 satisfies the so called Armijo condition, which is the exit condition in the while
 258 loop of Algorithm 3.2. Namely, $f(z^{k,i} + \alpha d^{k,i}) \leq f(z^{k,i}) + \gamma \alpha \nabla f(z^{k,i})^T d^{k,i}$ for all
 259 $\alpha \in [0, 2(1-\gamma)/L_{p_i^k, j(k)}]$ (see the proof of Proposition 3 in [12]). Since, in our line
 260 search, α is multiplied by $\delta \in (0, 1)$ until the Armijo condition is satisfied (see line 3
 261 in Algorithm 3.2), we immediately have the following result.

262 LEMMA 5.1. *At every inner iteration (k, i) , the initial stepsize $\Delta^{k,i}$ used in the*
 263 *Armijo line search is such that*

$$264 \quad \Delta^{k,i} \leq \frac{2(1-\gamma)}{L_{p_i^k, j(k)}} \Rightarrow \alpha^{k,i} = \Delta^{k,i},$$

$$265 \quad \Delta^{k,i} > \frac{2(1-\gamma)}{L_{p_i^k, j(k)}} \Rightarrow \alpha^{k,i} \in \left(\frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}}, \Delta^{k,i} \right],$$

267 *where, in the Armijo line search, $\gamma \in (0, 1)$ is the parameter for sufficient decrease*
 268 *and $\delta \in (0, 1)$ is the reduction parameter. Therefore, $\alpha^{k,i} \geq \min \left\{ \Delta^{k,i}, \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \right\}$.*

269 As a consequence of Lemma 1 in [12], we also have the following relation between
 270 the limit of $\{x^k\}$ and the limit of the sequences $\{z^{k,i}\}$, $i = 1, \dots, n$:

$$271 \quad (5.2) \quad \lim_{k \rightarrow \infty} x^k = x^* \Leftrightarrow \lim_{k \rightarrow \infty} z^{k,i} = x^*, \quad i = 1, \dots, n.$$

272 Now we state some useful properties derived from the semidefinite inner product
 273 and the seminorm defined at the end of Section 2. In the following results, we use L_j
 274 as defined in (2.5). The proofs are reported in Appendix A.

275 LEMMA 5.2. *For any $j \in \{1, \dots, n\}$ we have that*

$$276 \quad v^T(x' - x'') = \langle v - v_j e, x' - x'' \rangle_j, \quad \forall x', x'' \in \mathcal{F}, \quad \forall v \in \mathbb{R}^n.$$

277 LEMMA 5.3. *If f is convex over \mathbb{R}^n , for any $j \in \{1, \dots, n\}$ we have that*

$$278 \quad \left\| [\nabla f(x') - \nabla_j f(x')e] - [\nabla f(x'') - \nabla_j f(x'')e] \right\|_{(j)} \leq L_j \|x' - x''\|_{(j)}, \quad \forall x', x'' \in \mathcal{F}$$

279 COROLLARY 5.4. *If f is convex over \mathbb{R}^n , at every inner iteration (k, i) of AC2CD*
 280 *we have that*

$$281 \quad |\nabla_{p_i^k} f(v) - \nabla_{j(k)} f(v) + g^{k,i}| \leq L_{j(k)} \|v - z^{k,i}\|_{(j(k))}, \quad \forall v \in \mathbb{R}^n.$$

282 LEMMA 5.5. *If f is convex over \mathbb{R}^n , for any $j \in \{1, \dots, n\}$ we have that*

$$283 \quad f(x'') \leq f(x') + \nabla f(x')^T(x'' - x') + \frac{L_j}{2} \|x' - x''\|_{(j)}^2, \quad \forall x', x'' \in \mathcal{F}.$$

284 **6. Active-set identification in the non-convex case.** In this section, we
 285 show that AC2CD identifies the active set of problem (2.1) in a finite number of
 286 iterations, without any assumption on the convexity of f .

287 First of all, let us give the definition of active set for our problem.

288 DEFINITION 6.1. *Given a stationary point x^* of problem (2.1), we define the ac-*
 289 *tive set as*

$$290 \quad \mathcal{Z}(x^*) = \{i: x_i^* = l_i\} \cup \{i: x_i^* = u_i\}.$$

291 We also define

$$292 \quad \mathcal{Z}^+(x^*) = \mathcal{Z}(x^*) \cap \{i: \nabla_i f(x^*) \neq \lambda^*\},$$

293 where λ^* is the KKT multiplier associated with x^* appearing in (2.7).

We see that $\mathcal{L}(x^*)$ is the set of indices of all the variables that are at the lower or the upper bound in a stationary point x^* , whereas $\mathcal{L}^+(x^*)$ contains only the indices of the variables satisfying the strict complementarity. We notice that, from a geometric perspective, $\mathcal{L}^+(x^*)$ defines the face of \mathcal{F} exposed to $-\nabla f(x^*)$ [9].

The scope of this section is two-fold:

- (i) Firstly, it will be shown that, given a sequence of points $\{x^k\} \rightarrow x^*$ produced by AC2CD, an iteration \bar{k} exists such that, for all $k > \bar{k}$,

$$(6.1) \quad x_h^k = x_h^*, \quad \forall h \in \mathcal{L}^+(x^*).$$

Namely, in a finite number of iterations AC2CD sets to the bounds all the variables that satisfy the strict complementarity at x^* .

- (ii) Secondly, we will give a characterization of the neighborhood of x^* where (6.1) holds, which will be used in Section 7 to obtain an upper bound for \bar{k} (under convexity of f and a quadratic growth condition).

Note that, as common when analyzing active-set identification properties of an optimization algorithm, here we require the whole sequence $\{x^k\}$ to converge. For AC2CD, in [12] it was shown that every limit point of $\{x^k\}$ is stationary and, if $\{f(x^k)\}$ converges, then $\lim_{k \rightarrow \infty} \|z^{k,i+1} - z^{k,i}\| = 0$, $i = 1, \dots, n$, implying that $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ if a limit point of $\{x^k\}$ exists. So, using the same arguments given in [45, Theorem 14.1.5], we get that the whole sequence $\{x^k\}$ converges if the number of stationary points in \mathcal{L}^0 is finite. By a more general result stated in [21, Proposition 8.3.10], we also have that the whole sequence $\{x^k\}$ converges if it has an isolated limit point. Other conditions can be obtained from [5, Theorem 4.3]: if f satisfies a suitable descent property along the search directions, then a strict local minimum with no other stationary points in its neighborhood attracts the whole sequence $\{x^k\}$.

Now, we start our analysis by giving an intermediate result stating that, in a neighborhood of x^* , the index $j(k)$ is such that $l_{j(k)} < x_{j(k)}^* < u_{j(k)}$.

PROPOSITION 6.2. *Let $\{x^k\}$ be a sequence of points produced by AC2CD and assume that $\lim_{k \rightarrow \infty} x^k = x^*$. Define the maximum distance from the bounds at x^* as*

$$D^{\max}(x^*) = \max_{i=1, \dots, n} D_i(x^*),$$

which is positive by Assumption 1, and let k^j be the first outer iteration such that

$$\|x^k - x^*\|_\infty < \frac{\tau}{\tau + 1} D^{\max}(x^*), \quad \forall k \geq k^j,$$

where $\tau \in (0, 1]$ is the parameter used to choose $j(k)$, satisfying (3.3). Then, for all $k \geq k^j$ we have that $j(k) \notin \mathcal{L}(x^*)$.

Proof. Consider an outer iteration $k \geq k^j$ and let \hat{j} be an index such that $D_{\hat{j}}(x^*) = D^{\max}(x^*)$. We have $|x_{\hat{j}}^k - x_{\hat{j}}^*| \leq \|x^k - x^*\|_\infty < \frac{\tau}{\tau + 1} D^{\max}(x^*)$, implying that

$$(6.2) \quad x_{\hat{j}}^k - l_{\hat{j}} > x_{\hat{j}}^* - l_{\hat{j}} - \frac{\tau}{\tau + 1} D_{\hat{j}}(x^*) \quad \text{and} \quad u_{\hat{j}} - x_{\hat{j}}^k > u_{\hat{j}} - x_{\hat{j}}^* - \frac{\tau}{\tau + 1} D_{\hat{j}}(x^*).$$

Therefore, we can write

$$(6.3) \quad \begin{aligned} D_{\hat{j}}(x^k) &= \min\{x_{\hat{j}}^k - l_{\hat{j}}, u_{\hat{j}} - x_{\hat{j}}^k\} \stackrel{(6.2)}{>} \min\{x_{\hat{j}}^* - l_{\hat{j}}, u_{\hat{j}} - x_{\hat{j}}^*\} - \frac{\tau}{\tau + 1} D_{\hat{j}}(x^*) \\ &= D_{\hat{j}}(x^*) - \frac{\tau}{\tau + 1} D_{\hat{j}}(x^*) = \frac{1}{\tau + 1} D_{\hat{j}}(x^*). \end{aligned}$$

334 Arguing by contradiction, assume now that $j(k) \in \mathcal{Z}(x^*)$, that is,

$$335 \quad (6.4) \quad x_{j(k)}^* \in \{l_{j(k)}, u_{j(k)}\}.$$

336 We obtain

$$\begin{aligned} 337 \quad D_{j(k)}(x^k) &= \min\{x_{j(k)}^k - l_{j(k)}, u_{j(k)} - x_{j(k)}^k\} \stackrel{(6.4)}{\leq} |x_{j(k)}^k - x_{j(k)}^*| \leq \|x^k - x^*\|_\infty \\ &< \frac{\tau}{\tau+1} D^{\max}(x^*) = \frac{\tau}{\tau+1} D_j(x^*) \stackrel{(6.3)}{<} \tau D_j(x^k) \stackrel{(3.2)}{\leq} \tau D^k, \end{aligned}$$

338 contradicting (3.3). \square

339 Combining the above proposition with (5.2), the next result immediately follows.

340 PROPOSITION 6.3. *Let $\{x^k\}$ be a sequence of points produced by AC2CD and as-*
 341 *sume that $\lim_{k \rightarrow \infty} x^k = x^*$. There exists an iteration k^z such that, for all $k \geq k^z$,*

$$342 \quad l_{j(k)} < z_{j(k)}^{k,i} < u_{j(k)}, \quad i = 1, \dots, n+1.$$

343 Now, we are ready to show that (6.1) holds for all sufficiently large iterations.
 344 Our analysis takes inspiration from the one in [43] for proximal gradient methods,
 345 where it is proved that the active set is identified in a neighborhood of the optimal
 346 solution under the non-degeneracy assumption. That neighborhood is defined in [43]
 347 by using a problem-dependent constant related on “the amount of degeneracy” of the
 348 optimal solution.

349 Here, for a stationary point x^* such that $\mathcal{Z}^+(x^*) \neq \emptyset$, we define the following
 350 positive constant, measuring the “minimum amount of strict complementarity” at x^* :

$$351 \quad (6.5) \quad \zeta(x^*) = \min_{i \in \mathcal{Z}^+(x^*)} |\nabla_i f(x^*) - \lambda^*|,$$

352 where λ^* is the KKT multiplier associated to x^* , according to (2.7).

353 THEOREM 6.4. *Let $\{x^k\}$ be a sequence of points produced by AC2CD and assume*
 354 *that $\lim_{k \rightarrow \infty} x^k = x^*$. Let k be the first outer iteration such that*

$$355 \quad (6.6) \quad \|z^{k,i} - x^*\| < \frac{\zeta(x^*)}{2L + \max\left\{\frac{1}{A_l}, \frac{L^{\max}}{2(1-\gamma)}\right\}}, \quad i = 1, \dots, n, \quad \forall k \geq \bar{k},$$

356 where $\zeta(x^*) > 0$ is the minimum strict complementarity measure at x^* , defined as
 357 in (6.5), L is the Lipschitz constant of ∇f , $A_l > 0$ is the lower bound on the parameter
 358 $A^{k,i}$ used to compute the initial stepsize $\Delta^{k,i}$ in the Armijo line search (see (3.6)
 359 and (3.8)), $\gamma \in (0, 1)$ is the parameter for sufficient decrease in the Armijo line search
 360 and $L^{\max} > 0$ is the maximum among the local Lipschitz constants $L_{i,j}$, defined as
 361 in (2.4).

362 Also assume that $\bar{k} \geq \max\{k^j, k^z\}$, where k^j is the first outer iteration such that
 363 $j(k) \notin \mathcal{Z}(x^*)$ for all $k \geq k^j$, defined as in Proposition 6.2, and k^z is the first outer
 364 iteration such that $l_{j(k)} < z_{j(k)}^{k,i} < u_{j(k)}$, $i = 1, \dots, n+1$, for all $k \geq k^z$, defined as in
 365 Proposition 6.3.

366 Then, for all $k > \bar{k}$ we have that

$$367 \quad x_h^k = x_h^*, \quad \forall h \in \mathcal{Z}^+(x^*).$$

368 *Proof.* Consider an outer iteration $k \geq \bar{k}$ and any index $h \in \mathcal{Z}^+(x^*)$. Moreover,
 369 let (k, i) be the inner iteration where $p_i^k = h$. Without loss of generality, let us assume
 370 that $x_h^* = l_h$ (the proof for the case where $x_h^* = u_h$ is analogous). Namely,

$$371 \quad (6.7) \quad x_h^* = l_h \quad \text{and} \quad \nabla_h f(x^*) > \lambda^*,$$

372 where λ^* is the KKT multiplier associated to x^* , according to the stationary condi-
 373 tions (2.7). Since $k \geq k^j$, from Proposition 6.2 we have that

$$374 \quad (6.8) \quad j(k) \notin \mathcal{Z}(x^*),$$

375 implying that $h \neq j(k)$. Then, using (6.8) and the stationary conditions (2.7), we get
 376 $\lambda^* = \nabla_{j(k)} f(x^*)$. Recalling the definition of $\zeta(x^*)$, it follows that

$$377 \quad \zeta(x^*) \leq \nabla_h f(x^*) - \nabla_{j(k)} f(x^*).$$

378 Moreover, from the definition of $g^{k,i}$ given in (3.4) we can write

$$\begin{aligned} \nabla_h f(x^*) - \nabla_{j(k)} f(x^*) + g^{k,i} &= \nabla_h f(x^*) - \nabla_{j(k)} f(x^*) + \nabla_{j(k)} f(z^{k,i}) - \nabla_h f(z^{k,i}) \\ 379 \quad &\leq |\nabla_h f(x^*) - \nabla_h f(z^{k,i})| + |\nabla_{j(k)} f(z^{k,i}) - \nabla_{j(k)} f(x^*)| \\ &\leq 2\|\nabla f(x^*) - \nabla f(z^{k,i})\| \leq 2L\|x^* - z^{k,i}\|, \end{aligned}$$

380 and then,

$$381 \quad (6.9) \quad \zeta(x^*) \leq -g^{k,i} + 2L\|x^* - z^{k,i}\|.$$

382 Now, we can rewrite (6.6) by multiplying the numerator and the denominator of the
 383 right-hand side by $\max\left\{\frac{1}{A_l}, \frac{L^{\max}}{2(1-\gamma)}\right\}^{-1} = \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\}$, obtaining

$$384 \quad \|z^{k,i} - x^*\| < \frac{\zeta(x^*) \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\}}{2L \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\} + 1}.$$

385 Multiplying both sides of this inequality by the denominator of the right-hand side,
 386 we can write

$$\begin{aligned} \|z^{k,i} - x^*\| &= (\zeta(x^*) - 2L\|z^{k,i} - x^*\|) \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\} \\ 387 \quad (6.10) \quad &\stackrel{(6.9)}{\leq} -g^{k,i} \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\}. \end{aligned}$$

388 It follows that $g^{k,i} \leq 0$. If $g^{k,i} = 0$, we have

$$389 \quad x_h^{k+1} \stackrel{(5.1)}{=} z_h^{k,i+1} \stackrel{(3.4)}{=} z_h^{k,i} \stackrel{(6.10)}{=} x_h^*,$$

390 and the desired result is thus obtained. Now assume that $g^{k,i} < 0$. We can upper
 391 bound the largest feasible stepsize $\bar{\alpha}^{k,i}$ as follows:

$$\begin{aligned} \bar{\alpha}^{k,i} &\stackrel{(3.7)}{\leq} -\frac{z_h^{k,i} - l_h}{g^{k,i}} \stackrel{(6.7)}{=} -\frac{z_h^{k,i} - x_h^*}{g^{k,i}} \stackrel{(6.10)}{<} \min\left\{A_l, \frac{2(1-\gamma)}{L^{\max}}\right\} \\ 392 \quad (6.11) \quad &\stackrel{(3.8)}{\leq} \min\left\{A^{k,i}, \frac{2(1-\gamma)}{L^{\max}}\right\}, \end{aligned}$$

393 implying that $\bar{\alpha}^{k,i} < A^{k,i}$. Taking into account that the initial stepsize $\Delta^{k,i}$ in
 394 the Armijo line search is chosen as in (3.6), we have that $\Delta^{k,i} = \bar{\alpha}^{k,i}$. So, using
 395 again (6.11) we obtain

$$396 \quad \Delta^{k,i} = \bar{\alpha}^{k,i} < \frac{2(1-\gamma)}{L^{\max}} \leq \frac{2(1-\gamma)}{L_{h,j(k)}}.$$

397 From Lemma 5.1 we get that $\alpha^{k,i} = \bar{\alpha}^{k,i}$. Since $\bar{\alpha}^{k,i}$ is the largest feasible stepsize
 398 along $d^{k,i}$, (at least) one variable between $z_h^{k,i+1}$ and $z_{j(k)}^{k,i+1}$ will be at the lower or
 399 the upper bound. Using the fact that $k \geq k^z$, from Proposition 6.3 we have that
 400 $z_{j(k)}^{k,i+1} \in (l_{j(k)}, u_{j(k)})$, and then $z_h^{k,i+1}$ will be necessarily at the lower or the upper
 401 bound. Since $g^{k,i} < 0$, from the definition of the search direction given in (3.4) it
 402 follows that $z_h^{k,i+1} = l_h$. Using (5.1) and (6.7), we finally have that $z_h^{k,i+1} = x_h^{k+1}$ and
 403 $l_h = x_h^*$, yielding to the desired result. \square

404 *Remark 6.5.* From (5.2), there must exist an outer iteration \bar{k} such that (6.6)
 405 holds, provided the whole sequence $\{x^k\}$ converges to x^* and $\mathcal{Z}^+(x^*) \neq \emptyset$.

406 **7. Active-set complexity.** In this section, the main result of the paper is pre-
 407 sented: under convexity of f and a quadratic growth condition (satisfied by any
 408 strongly convex function), it is possible to compute the maximum number of itera-
 409 tions required by AC2CD to identify the active set, thus extending what obtained
 410 in the previous section. Using the definition given in [43], we refer to the maximum
 411 number of iterations required to identify the active set as “active-set complexity”.

412 To obtain the desired result, we first show how choosing the initial stepsize in
 413 the Armijo line search, in order to meet an additional requirement. Then, we will
 414 show non-asymptotic sublinear convergence rate of AC2CD, which, combined with
 415 Theorem 6.4, will lead to the active-set complexity of the algorithm.

416 **7.1. Initial stepsize in the Armijo line search.** To obtain non-asymptotic
 417 sublinear convergence rate of AC2CD, for all $k \geq 0$ we need to satisfy

$$418 \quad (7.1) \quad l_{j(k)} < z_{j(k)}^{k,i} < u_{j(k)}, \quad i = 1, \dots, n+1.$$

419 Note that, in general, (7.1) holds only for sufficiently large k (see the proof of
 420 Theorem 1 in [12]). To satisfy (7.1) for all $k \geq 0$ we can use sufficiently small stepsizes
 421 in all the inner iterations, exploiting the fact that $x_{j(k)}^k = z_{j(k)}^{k,1} \in (l_{j(k)}, u_{j(k)})$ for all
 422 $k \geq 0$. In particular, to obtain a small stepsize $\alpha^{k,i}$ from the Armijo line search we
 423 must choose a small value of the initial stepsize $\Delta^{k,i}$. Taking into account (3.6), this
 424 means that we must use a small value of $A^{k,i}$. Anyway, we have to keep in mind that
 425 $A^{k,i}$ must satisfy (3.8) as well. A possible strategy is setting $A_u > 0$, $\epsilon \in (0, 1)$ and,
 426 at every inner iteration (k, i) , computing

$$427 \quad (7.2) \quad A^{k,i} = \begin{cases} \min\{\hat{\alpha}^{k,i}, A_u\}, & \text{if } g^{k,i} \neq 0 \text{ (i.e., if } d^{k,i} \text{ is a non-zero direction),} \\ A_u, & \text{otherwise,} \end{cases}$$

428 where $\hat{\alpha}^{k,i}$ is the stepsize such that $D_{j(k)}(z^{k,i} + \hat{\alpha}^{k,i} d^{k,i}) = \epsilon D_{j(k)}(z^{k,i})$ when $g^{k,i} \neq 0$.
 429 Note that $\hat{\alpha}^{k,i}$ may be infeasible and/or infinity. Since $\alpha^{k,i} \leq A^{k,i} \leq \hat{\alpha}^{k,i}$, it follows
 430 that $D_{j(k)}(z^{k,i+1}) = D_{j(k)}(z^{k,i} + \alpha^{k,i} d^{k,i}) \geq D_{j(k)}(z^{k,i} + \hat{\alpha}^{k,i} d^{k,i}) \geq \epsilon D_{j(k)}(z^{k,i})$.
 431 Consequently,

$$432 \quad (7.3) \quad D_{j(k)}(z^{k,i+1}) \geq \epsilon^i D_{j(k)}(z^{k,1}) = \epsilon^i D_{j(k)}(x^k) > 0, \quad i = 1, \dots, n.$$

433 Then, this choice of $A^{k,i}$ satisfies (7.1) for all $k \geq 0$. To show that it also satisfies (3.8),
 434 we have to explicitly write the expression of $\hat{\alpha}^{k,i}$, which can be obtained by simple
 435 calculations (recall that $\hat{\alpha}^{k,i}$ is defined only when $g^{k,i} \neq 0$):

436 If $g^{k,i} > 0$,

$$437 \quad \hat{\alpha}^{k,i} = \begin{cases} (1 - \epsilon)D_{j(k)}(z^{k,i})/g^{k,i}, & \text{if } D_{j(k)}(z^{k,i}) = z_{j(k)}^{k,i} - l_{j(k)}, \\ [z_{j(k)}^{k,i} - l_{j(k)} - \epsilon D_{j(k)}(z^{k,i})]/g^{k,i}, & \text{otherwise;} \end{cases}$$

438 else if $g^{k,i} < 0$,

$$439 \quad \hat{\alpha}^{k,i} = \begin{cases} (1 - \epsilon)D_{j(k)}(z^{k,i})/|g^{k,i}|, & \text{if } D_{j(k)}(z^{k,i}) = u_{j(k)} - z_{j(k)}^{k,i}, \\ [u_{j(k)} - z_{j(k)}^{k,i} - \epsilon D_{j(k)}(z^{k,i})]/|g^{k,i}|, & \text{otherwise.} \end{cases}$$

441 We see that, when $g^{k,i} \neq 0$, we have $\hat{\alpha}^{k,i} \geq (1 - \epsilon)D_{j(k)}(z^{k,i})/|g^{k,i}|$. Using (7.2)
 442 and (7.3), it follows that

$$443 \quad \min \left\{ \frac{(1 - \epsilon)\epsilon^{i-1}D_{j(k)}(x^k)}{|g^{k,i}|}, A_u \right\} \leq A^{k,i} \leq A_u, \quad \text{if } g^{k,i} \neq 0.$$

444 Then, (3.8) is satisfied with a proper value of A_l which can be easily obtained, since
 445 any non-zero $|g^{k,i}|$ is less than or equal to $\max_{i,j=1,\dots,n} \{\nabla_j f(x) - \nabla_i f(x) : x \in \mathcal{L}^0\}$
 446 (which is finite by the assumption that the level set \mathcal{L}^0 is compact) and, from (3.3),
 447 we have $D_{j(k)}(x^k) \geq \tau \min_{x \in \mathcal{L}^0} \max_{i=1,\dots,n} D_i(x)$ (which is positive by Assumption 1).

448 Many other strategies can be used to compute a value of $A^{k,i}$ that satisfies all
 449 the required conditions. It is important to note that, in practice, $A^{k,i}$ should not be
 450 too small compared to the largest feasible stepsize $\bar{\alpha}^{k,i}$ (for a non-zero direction $d^{k,i}$),
 451 otherwise the Armijo line search may produce extremely small stepsizes which can
 452 dramatically slow down the algorithm. For example, ϵ should be sufficiently smaller
 453 than 1 in the above described strategy.

454 In the rest of this section, we will assume $A^{k,i}$ to be computed in order to satisfy,
 455 together with (3.8), condition (7.1) for all $k \geq 0$.

456 **7.2. Convergence rate analysis.** In this subsection we show that, when f is
 457 convex, AC2CD has a non-asymptotic sublinear convergence rate. Let us remark that
 458 the results reported here are completely different from those given in [12], where a
 459 linear rate was obtained, but asymptotically, whereas a non-asymptotic linear rate
 460 was proved only when there are no bounds on the variables (both results are not
 461 useful in the analysis of the active-set complexity).

462 Our results here are obtained by adapting the analysis of the block coordinate
 463 gradient projection method in [2] for minimization problems over the Cartesian prod-
 464 uct of closed convex sets. In particular, with respect to [2], major difficulties in our
 465 analysis come from the presence of the coupling constraint in the problem and the
 466 absence of projection operations in the algorithm. In such a context, the next lemma
 467 establishes a useful property of AC2CD.

468 **LEMMA 7.1.** *For all $x^* \in X^*$, at every inner iteration (k, i) of AC2CD we have
 469 that*

$$470 \quad g^{k,i}(x_{p_i^*}^* - z_{p_i^*}^{k,i+1}) \leq \max \left\{ \frac{1}{A_l}, \frac{L^{max}}{2\delta(1 - \gamma)} \right\} |z_{p_i^*}^{k,i+1} - z_{p_i^*}^{k,i}| |x_{p_i^*}^* - z_{p_i^*}^{k,i+1}|,$$

471 where $A_l > 0$ is the lower bound on the parameter $A^{k,i}$ used to compute the initial
 472 stepsize $\Delta^{k,i}$ in the Armijo line search (see (3.6) and (3.8)), $\gamma \in (0, 1)$ is the parameter

473 for sufficient decrease in the Armijo line search, $\delta \in (0, 1)$ is the reduction parameter
 474 in the Armijo line search and $L^{\max} > 0$ is the maximum among the local Lipschitz
 475 constants $L_{i,j}$, defined in (2.4).

476 *Proof.* Consider any inner iteration (k, i) . The result is trivial if $g^{k,i} = 0$, so we
 477 assume that $g^{k,i} \neq 0$ and distinguish two possible cases.

- 478 (i) First, assume that $\alpha^{k,i} = \bar{\alpha}^{k,i}$, that is, the largest feasible stepsize is used. This
 479 means that (at least) one variable between $z_{p_i^k}^{k,i+1}$ and $z_{j(k)}^{k,i+1}$ will be at the lower
 480 or the upper bound. Recalling that (7.1) holds for all $k \geq 0$, necessarily $z_{p_i^k}^{k,i+1}$
 481 will be at the lower or the upper bound. Using the definition of $\bar{\alpha}^{k,i}$ given
 482 in (3.7), it follows that either $z_{p_i^k}^{k,i+1} = u_{p_i^k}$ if $g^{k,i} > 0$, or $z_{p_i^k}^{k,i+1} = l_{p_i^k}$ if $g^{k,i} < 0$,
 483 implying that $g^{k,i}(x_{p_i^k}^* - z_{p_i^k}^{k,i+1}) \leq 0$ and the desired result is obtained.
 484 (ii) Now, assume that $\alpha^{k,i} < \bar{\alpha}^{k,i}$, which implies that $\bar{\alpha}^{k,i} > 0$ and, from (3.9), that
 485 $z^{k,i+1} \neq z^{k,i}$. Since $z_{p_i^k}^{k,i+1} = z_{p_i^k}^{k,i} + \alpha^{k,i} g^{k,i}$, it follows that $\alpha^{k,i} g^{k,i} \neq 0$. Recalling
 486 the definition of $g^{k,i}$ given in (3.4), this implies that $p_i^k \neq j(k)$. Moreover, we
 487 can write

$$488 \quad g^{k,i}(x_{p_i^k}^* - z_{p_i^k}^{k,i+1}) = \frac{(z_{p_i^k}^{k,i+1} - z_{p_i^k}^{k,i})(x_{p_i^k}^* - z_{p_i^k}^{k,i+1})}{\alpha^{k,i}} \leq \frac{|z_{p_i^k}^{k,i+1} - z_{p_i^k}^{k,i}| |x_{p_i^k}^* - z_{p_i^k}^{k,i+1}|}{\alpha^{k,i}}.$$

489 So, to obtain the desired result we have to show that

$$490 \quad \alpha^{k,i} \geq \min \left\{ A_l, \frac{2\delta(1-\gamma)}{L^{\max}} \right\}.$$

491 To this extent, let us distinguish two further subcases, depending on whether
 492 $\Delta^{k,i} = \bar{\alpha}^{k,i}$ or $\Delta^{k,i} < \bar{\alpha}^{k,i}$, according to the definition of $\Delta^{k,i}$ given in (3.6).

- 493 • If $\Delta^{k,i} = \bar{\alpha}^{k,i}$, then $\alpha^{k,i} < \Delta^{k,i}$ (recall that we are considering the case
 494 $\alpha^{k,i} < \bar{\alpha}^{k,i}$) and, from Lemma 5.1, it follows that

$$495 \quad \alpha^{k,i} > \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \geq \min \left\{ A_l, \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \right\} \geq \min \left\{ A_l, \frac{2\delta(1-\gamma)}{L^{\max}} \right\}.$$

- 496 • If $\Delta^{k,i} < \bar{\alpha}^{k,i}$, from (3.6) we have $\Delta^{k,i} = A^{k,i}$. Using Lemma 5.1 it follows
 497 that

$$498 \quad \alpha^{k,i} \geq \min \left\{ \Delta^{k,i}, \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \right\} = \min \left\{ A^{k,i}, \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \right\}$$

$$499 \quad \geq \min \left\{ A_l, \frac{2\delta(1-\gamma)}{L_{p_i^k, j(k)}} \right\} \geq \min \left\{ A_l, \frac{2\delta(1-\gamma)}{L^{\max}} \right\}. \quad \square$$

501 Now, we give a first result on the decrease in the objective function at every outer
 502 iteration.

503 **PROPOSITION 7.2.** *At every outer iteration k of AC2CD we have that*

$$504 \quad f(x^k) - f(x^{k+1}) \geq \frac{\gamma}{A_u} \|x^{k+1} - x^k\|_{j(k)}^2,$$

505 where $A_u > 0$ is the upper bound on the parameter $A^{k,i}$ used to compute the initial
 506 stepsize $\Delta^{k,i}$ in the Armijo line search (see (3.6) and (3.8)) and $\gamma \in (0, 1)$ is the
 507 parameter for sufficient decrease in the Armijo line search.

508 *Proof.* First we show that at, every inner iteration (k, i) , we have

$$509 \quad (7.4) \quad f(z^{k,i}) - f(z^{k,i+1}) \geq \frac{\gamma}{A_u} (z_{p_i^k}^{k,i+1} - z_{p_i^k}^{k,i})^2.$$

510 If $\alpha^{k,i} = 0$, then $z^{k,i+1} = z^{k,i}$ and (7.4) trivially holds. If $\alpha^{k,i} > 0$, from the instruc-
511 tions of the Armijo line search it follows that $f(z^{k,i+1}) \leq f(z^{k,i}) + \gamma \alpha^{k,i} \nabla f(z^{k,i})^T d^{k,i}$.
512 Using (3.5), we can write

$$513 \quad f(z^{k,i+1}) \leq f(z^{k,i}) - \gamma \alpha^{k,i} (g^{k,i})^2 = f(z^{k,i}) - \frac{\gamma}{\alpha^{k,i}} (\alpha^{k,i} g^{k,i})^2.$$

514 Since $z_{p_i^k}^{k,i+1} = z^{k,i} + \alpha^{k,i} g^{k,i}$ and $\alpha^{k,i} \leq A_u$, we obtain (7.4). Hence, we have

$$\begin{aligned} 515 \quad f(x^k) - f(x^{k+1}) &= \sum_{i=1}^n [f(z^{k,i}) - f(z^{k,i+1})] \stackrel{(7.4)}{\geq} \frac{\gamma}{A_u} \sum_{i=1}^n (z_{p_i^k}^{k,i+1} - z_{p_i^k}^{k,i})^2 \\ 516 \quad &= \frac{\gamma}{A_u} \sum_{i: p_i^k \neq j(k)} (z_{p_i^k}^{k,i+1} - z_{p_i^k}^{k,i})^2 \stackrel{(5.1)}{=} \frac{\gamma}{A_u} \sum_{i: p_i^k \neq j(k)} (x_{p_i^k}^{k+1} - x_{p_i^k}^k)^2 \\ 517 \quad &= \frac{\gamma}{A_u} \|x^{k+1} - x^k\|_{\langle j(k) \rangle}^2, \\ 518 \end{aligned}$$

519 where, in the second equality, we have used the fact that $z_{j(k)}^{k,i+1} = z_{j(k)}^{k,i}$ when $p_i^k = j(k)$,
520 according to the definition of the search direction $d^{k,i}$ given in (3.4). \square

521 In the rest of this section, the objective function will be required to be convex
522 over \mathbb{R}^n and its optimal value for problem (2.1) will be denoted by f^* . Let us also
523 define the following constants (which are finite under convexity of f , since this implies
524 $X^* \subseteq \mathcal{L}^0$, where the level set \mathcal{L}^0 is assumed to be non-empty and compact):

$$525 \quad (7.5) \quad R^0 = \max_{\substack{j=1, \dots, n \\ x \in \mathcal{L}^0 \\ x^* \in X^*}} \|x - x^*\|_{\langle j \rangle},$$

$$526 \quad (7.6) \quad G^* = \max_{\substack{i,j=1, \dots, n \\ x^* \in X^*}} [\nabla_j f(x^*) - \nabla_i f(x^*)].$$

528 We see that R^0 is the maximum distance between a point in the level set \mathcal{L}^0 and a
529 point in X^* , where the distance is measured in terms of the pseudometrics induced
530 by the seminorms $\|\cdot\|_{\langle j \rangle}$ (the latter can be upper bounded by the Euclidean norm,
531 see (2.11)). From the KKT conditions (2.7), we also note that G^* is related to the
532 minimum strict complementarity measure $\zeta(x^*)$ defined in (6.5), in the sense that, if
533 $\mathcal{L}^+(x^*) \neq \emptyset$ for some $x^* \in X^*$, then $G^* \geq \zeta(x^*) > 0$, while, if $\mathcal{L}^+(x^*) = \emptyset$ for all
534 $x^* \in X^*$, then $G^* = 0$ and $\zeta(x^*)$ is not defined for any $x^* \in X^*$. We can interpret G^*
535 as a measure of the ‘‘maximum amount of strict complementarity’’ over the set X^* .

536 We now state a result which, for every outer iteration, relates the decrease in the
537 objective function with the optimization error.

538 **PROPOSITION 7.3.** *Assume that f is convex over \mathbb{R}^n . Then, at every outer iter-*
539 *ation k of AC2CD we have that*

$$540 \quad f(x^k) - f(x^{k+1}) \geq \frac{\gamma(f(x^{k+1}) - f^*)^2}{A_u(n-1) \left[\left(\max \left\{ \frac{1}{A_l}, \frac{L^{max}}{2\delta(1-\gamma)} \right\} + 2\hat{L}^{max} \right) R^0 + G^* \right]^2},$$

541 where $A_l > 0$ and $A_u > 0$ are the lower and the upper bound, respectively, on the
 542 parameter $A^{k,i}$ used to compute the initial stepsize $\Delta^{k,i}$ in the Armijo line search
 543 (see (3.6) and (3.8)), $\delta \in (0, 1)$ is the reduction parameter in the Armijo line search,
 544 $\gamma \in (0, 1)$ is the parameter for sufficient decrease in the Armijo line search, $L^{\max} > 0$
 545 is the maximum among the local Lipschitz constants $L_{i,j}$, defined as in (2.4), $\hat{L}^{\max} > 0$
 546 is the maximum among the constants $L_j = \sum_{i=1}^n L_{i,j}$, defined as in (2.6), $R^0 \geq 0$ is
 547 the maximum distance between a point in the level set \mathcal{L}^0 and an optimal solution,
 548 defined as in (7.5), and $G^* \geq 0$ is the maximum strict complementarity measure over
 549 X^* , defined as in (7.6).

550 *Proof.* Let x^* be an optimal solution of problem (2.1) and consider any inner
 551 iteration (k, i) . From the definition of the search direction $d^{k,i}$ given in (3.4), we have
 552 that $z_{p_i^k}^{k,i+1} \geq z_{p_i^k}^{k,i}$ if $g^{k,i} \geq 0$, and $z_{p_i^k}^{k,i+1} \leq z_{p_i^k}^{k,i}$ if $g^{k,i} \leq 0$. Namely, $g^{k,i}(z_{p_i^k}^{k,i} - z_{p_i^k}^{k,i+1}) \leq$
 553 0 and, using (5.1), we can write $g^{k,i}(x_{p_i^k}^k - x_{p_i^k}^{k+1}) \leq 0$. Then,

$$\begin{aligned} g^{k,i}(x_{p_i^k}^k - x_{p_i^k}^*) &\leq g^{k,i}(x_{p_i^k}^{k+1} - x_{p_i^k}^*) \\ &= [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1})](x_{p_i^k}^* - x_{p_i^k}^{k+1}) \\ &\quad + [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1}) + g^{k,i}](x_{p_i^k}^{k+1} - x_{p_i^k}^*). \end{aligned}$$

555 Using Corollary 5.4 with $v = x^{k+1}$, we have that

$$\begin{aligned} \nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1}) + g^{k,i} &\leq \hat{L}^{\max} \|z^{k,i} - x^{k+1}\|_{\langle j(k) \rangle} \\ &\stackrel{(5.1)}{\leq} \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle}. \end{aligned}$$

557 It follows that

$$\begin{aligned} g^{k,i}(x_{p_i^k}^k - x_{p_i^k}^*) &\leq [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1})](x_{p_i^k}^* - x_{p_i^k}^{k+1}) \\ &\quad + \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle} |x_{p_i^k}^* - x_{p_i^k}^{k+1}|. \end{aligned}$$

559 Summing these inequalities, we obtain

$$\begin{aligned} \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^k - x_{p_i^k}^*) &\leq \sum_{i: p_i^k \neq j(k)} [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1})](x_{p_i^k}^* - x_{p_i^k}^{k+1}) \\ &\quad + \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle} \sum_{i: p_i^k \neq j(k)} |x_{p_i^k}^* - x_{p_i^k}^{k+1}| \\ &\stackrel{(2.10)}{\leq} \sum_{i: p_i^k \neq j(k)} [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1})](x_{p_i^k}^* - x_{p_i^k}^{k+1}) \\ &\quad + \sqrt{n-1} R^0 \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle}. \end{aligned}$$

561 Using Lemma 5.2 with $v = \nabla f(x^{k+1})$, $x' = x^*$ and $x'' = x^{k+1}$, we can write

$$\begin{aligned} \nabla f(x^{k+1})^T (x^* - x^{k+1}) &= \langle \nabla f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1}), x^* - x^{k+1} \rangle_{j(k)} \\ &= \sum_{i: p_i^k \neq j(k)} [\nabla_{p_i^k} f(x^{k+1}) - \nabla_{j(k)} f(x^{k+1})](x_{p_i^k}^* - x_{p_i^k}^{k+1}), \end{aligned}$$

563 and then,

$$564 \quad \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^k - x_{p_i^k}^*) \leq \nabla f(x^{k+1})^T(x^* - x^{k+1}) + \sqrt{n-1} R^0 \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle}.$$

565 From the convexity of f we have that $f(x^{k+1}) - f^* \leq \nabla f(x^{k+1})^T(x^{k+1} - x^*)$. Hence,

$$\begin{aligned} f(x^{k+1}) - f^* &\leq \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^* - x_{p_i^k}^k) + \sqrt{n-1} R^0 \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle} \\ 566 \quad &= \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^* - x_{p_i^k}^{k+1}) + \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^{k+1} - x_{p_i^k}^k) \\ &\quad + \sqrt{n-1} R^0 \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle}. \end{aligned}$$

567 Using (5.1) and Lemma 7.1, for all i such that $p_i^k \neq j(k)$ we can write

$$568 \quad g^{k,i}(x_{p_i^k}^* - x_{p_i^k}^{k+1}) \leq \max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2\delta(1-\gamma)} \right\} |x_{p_i^k}^{k+1} - x_{p_i^k}^k| |x_{p_i^k}^* - x_{p_i^k}^{k+1}|.$$

569 Therefore,

$$\begin{aligned} f(x^{k+1}) - f^* &\leq \max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2\delta(1-\gamma)} \right\} \sum_{i: p_i^k \neq j(k)} |x_{p_i^k}^{k+1} - x_{p_i^k}^k| |x_{p_i^k}^* - x_{p_i^k}^{k+1}| \\ 570 \quad (7.7) \quad &\quad + \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^{k+1} - x_{p_i^k}^k) \\ &\quad + \sqrt{n-1} R^0 \hat{L}^{\max} \|x^k - x^{k+1}\|_{\langle j(k) \rangle}. \end{aligned}$$

571 To obtain the desired result, now we upper bound the two summations in the right-
572 hand side of (7.7) by appropriate constants.

573 • As for the first summation in the right-hand side of (7.7), using the fact that
574 $|x_{p_i^k}^* - x_{p_i^k}^{k+1}| \leq \|x^* - x^{k+1}\|_{\langle j(k) \rangle}$ by (2.9), we can write

$$\begin{aligned} 575 \quad (7.8) \quad \sum_{i: p_i^k \neq j(k)} |x_{p_i^k}^{k+1} - x_{p_i^k}^k| |x_{p_i^k}^* - x_{p_i^k}^{k+1}| &\leq R^0 \sum_{i: p_i^k \neq j(k)} |x_{p_i^k}^{k+1} - x_{p_i^k}^k| \\ &\stackrel{(2.10)}{\leq} \sqrt{n-1} R^0 \|x^{k+1} - x^k\|_{\langle j(k) \rangle}. \end{aligned}$$

576 • As for the second summation in the right-hand side of (7.7), from the trian-
577 gular inequality we have that

$$578 \quad g^{k,i} \leq |g^{k,i} + \nabla_{p_i^k} f(x^*) - \nabla_{j(k)} f(x^*)| + |\nabla_{p_i^k} f(x^*) - \nabla_{j(k)} f(x^*)|,$$

579 and then, using Corollary 5.4 with $v = x^*$,

$$580 \quad g^{k,i} \leq L_{j(k)} \|x^* - z^{k,i}\|_{\langle j(k) \rangle} + |\nabla_{p_i^k} f(x^*) - \nabla_{j(k)} f(x^*)| \leq \hat{L}^{\max} R^0 + G^*.$$

581 Taking into account (2.10), we get

$$582 \quad (7.9) \quad \sum_{i: p_i^k \neq j(k)} g^{k,i}(x_{p_i^k}^{k+1} - x_{p_i^k}^k) \leq (\hat{L}^{\max} R^0 + G^*) \sqrt{n-1} \|x^{k+1} - x^k\|_{\langle j(k) \rangle}.$$

583 Combining (7.7) with (7.8) and (7.9), we have that

$$584 \quad f(x^{k+1}) - f^* \leq \sqrt{n-1} \left[\left(\max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2\delta(1-\gamma)} \right\} + 2\hat{L}^{\max} \right) R^0 + G^* \right] \|x^{k+1} - x^k\|_{\langle j(k) \rangle}.$$

585 Using Proposition 7.2, the desired result is finally obtained. \square

586 We are now ready to show the non-asymptotic sublinear convergence rate of
587 AC2CD.

588 **THEOREM 7.4.** *Assume that f is convex over \mathbb{R}^n . Then, at every outer iteration*
589 *$k \geq 1$ of AC2CD we have that*

$$590 \quad f(x^k) - f^* \leq \frac{C}{k},$$

591 where C is equal to

$$592 \quad \sqrt{n-1} \max \left\{ \frac{3A_u \sqrt{n-1}}{2\gamma}, \frac{1}{L^{\max}} \right\} \left[\left(\max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2\delta(1-\gamma)} \right\} + 2\hat{L}^{\max} \right) R^0 + 2G^* \right]^2,$$

593 $A_l > 0$ and $A_u > 0$ are the lower and the upper bound, respectively, on the parameter
594 $A^{k,i}$ used to compute the initial stepsize $\Delta^{k,i}$ in the Armijo line search (see (3.6)
595 and (3.8)), $\delta \in (0, 1)$ is the reduction parameter in the Armijo line search, $\gamma \in (0, 1)$
596 is the parameter for sufficient decrease in the Armijo line search, $L^{\max} > 0$ is the
597 maximum among the local Lipschitz constants $L_{i,j}$, defined as in (2.4), $\hat{L}^{\max} > 0$ is
598 the maximum among the constants $L_j = \sum_{i=1}^n L_{i,j}$, defined as in (2.6), $R^0 \geq 0$ is
599 the maximum distance between a point in the level set \mathcal{L}^0 and an optimal solution,
600 defined as in (7.5), and $G^* \geq 0$ is the maximum strict complementarity measure over
601 X^* , defined as in (7.6).

602 *Proof.* Consider a sequence $\{a^k\}$ of nonnegative scalars such that $a^k - a^{k+1} \geq$
603 $\beta(a^{k+1})^2$, for all $k \geq 0$, with $\beta > 0$. From Lemma 6.2 in [2] we have that, if $a^1 \leq$
604 $3/(2\beta)$ and $a^2 \leq 3/(4\beta)$, then $a^k \leq 3/(2\beta k)$, for all $k \geq 1$. Using $a^k = f(x^k) - f^*$,
605 in view of Proposition 7.3 we have that $a^k - a^{k+1} \geq \beta(a^{k+1})^2$ with $\beta \geq 3/(2C)$. It
606 follows that the desired result is obtained if

$$607 \quad (7.10) \quad f(x^1) - f^* \leq C \quad \text{and} \quad f(x^2) - f^* \leq \frac{C}{2}.$$

608 To show that (7.10) holds, by definition of C we first write

$$609 \quad (7.11) \quad \begin{aligned} C &\geq \frac{\sqrt{n-1}}{L^{\max}} \left[\left(\max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2\delta(1-\gamma)} \right\} + 2\hat{L}^{\max} \right) R^0 + 2G^* \right]^2 \\ &\geq \frac{\sqrt{n-1}}{L^{\max}} \left[\left(\frac{L^{\max}}{2} + 2\hat{L}^{\max} \right) R^0 + 2G^* \right]^2, \end{aligned}$$

610 where the last inequality follows from the fact that $2\delta(1-\gamma) \leq 2$, since $\delta, \gamma \in (0, 1)$.
611 Now, we use the trivial inequality $(\theta_1 + \theta_2 + \theta_3)^2 \geq 2\theta_1(\theta_2 + \theta_3)$, holding for all
612 $\theta_1, \theta_2, \theta_3 \in \mathbb{R}$, with the choice $\theta_1 = L^{\max} R^0/2, \theta_2 = 2\hat{L}^{\max} R^0, \theta_3 = 2G^*$. We get

$$613 \quad (7.12) \quad C \geq 2\sqrt{n-1} R^0 (\hat{L}^{\max} R^0 + G^*) \geq 2 \left[\frac{\hat{L}^{\max}}{2} (R^0)^2 + \sqrt{n-1} G^* R^0 \right],$$

614 where the last inequality follows from the fact that we are assuming $n \geq 2$.

615 Now consider an outer iteration $k \geq 1$, picking any $x^* \in X^*$ and any $j \in$
 616 $\{1, \dots, n\}$. Using Lemma 5.2 with $v = \nabla f(x^*)$, $x' = x^k$ and $x'' = x^*$, we have

$$\begin{aligned} \nabla f(x^*)^T(x^k - x^*) &= \langle \nabla f(x^*) - \nabla_j f(x^*)e, x^k - x^* \rangle_j \\ &= \sum_{h \neq j} [\nabla_h f(x^*) - \nabla_j f(x^*)](x_h^k - x_h^*) \\ &\stackrel{(2.10)}{\leq} \sqrt{n-1}G^* \|x^k - x^*\|_{(j)}. \end{aligned}$$

618 So, using Lemma 5.5 with $x' = x^*$ and $x'' = x^k$, we get

$$619 \quad f(x^k) - f^* \leq \nabla f(x^*)^T(x^k - x^*) + \frac{L_j}{2} \|x^* - x^k\|_{(j)}^2 \leq \sqrt{n-1}G^*R^0 + \frac{\hat{L}^{\max}}{2}(R^0)^2.$$

620 In view of (7.12), we conclude that $f(x^k) - f^* \leq C/2$, implying that (7.10) holds. \square

621 A question that can naturally arise is whether the constant C in Theorem 7.4
 622 is tight. To answer this challenging question, we can look in detail at the steps of
 623 the above proofs, from which it seems that C may in fact be loose. For example, in
 624 the proof of Theorem 7.4 we got a lower bound for C by decomposing the last term
 625 in (7.11) as the sum of $\hat{L}^{\max}(R^0)^2 + 2\sqrt{n-1}G^*R^0$ and

$$626 \quad (2\sqrt{n-1}-1)\hat{L}^{\max}(R^0)^2 + \sqrt{n-1} \left[\frac{L^{\max}(R^0)^2}{4} + 4 \frac{(\hat{L}^{\max}R^0)^2 + (G^*)^2 + 2\hat{L}^{\max}G^*R^0}{L^{\max}} \right].$$

627 We then obtained (7.12) by lower bounding the above quantity by 0. But the above
 628 quantity may be much larger than 0 and, for large values of n and G^* , even dominant
 629 over $\hat{L}^{\max}(R^0)^2 + 2\sqrt{n-1}G^*R^0$, observing that $\hat{L}^{\max} = \xi L^{\max}$, with $\xi \in [1, n-1]$,
 630 as we see from (2.4), (2.5) and (2.6).

631 In the literature, a non-asymptotic convergence rate was also shown for other
 632 coordinate descent methods on different settings with one or more linear constraints,
 633 where the working set is chosen by random selection [37, 38, 40, 46, 48] or by rules
 634 based on first-order optimality violation [1, 28]. In particular, just like AC2CD,
 635 random coordinate descent do not use ∇f to choose the working set. A sublinear
 636 rate (in expectation) with respect to the objective values was shown for random
 637 coordinate descent in [40] under convexity of f , and a linear rate (in expectation) was
 638 shown in [48] under the additional assumption of proximal-PL inequality. We note
 639 that the sublinear rate $f(x^k) - f^* \leq n^2 L(\bar{R}^0)^2 / [k + n^2 L(\bar{R}^0)^2 / (f(x^0) - f^*)]$ obtained
 640 for random coordinate descent in [40], where $\bar{R}^0 = \max_x \{ \max_{x^* \in X^*} \|x - x^*\| : x \in \mathcal{L}^0 \}$,
 641 holds with respect to the inner iterations, so k should be multiplied by a factor $\mathcal{O}(n)$
 642 to have a fair comparison with AC2CD, for which the rate was computed with respect
 643 to the outer iterations. With this adjustment, the rate of random coordinate descent
 644 is however better than $f(x^k) - f^* \leq C/k$ obtained for AC2CD, with the constant C
 645 from Theorem 7.4 being $\mathcal{O}(n(L^{\max}R^0 + G^*)^2)$ if we reasonably assume $\sqrt{n} \gg 1/L^{\max}$
 646 and consider $\hat{L}^{\max} = \mathcal{O}(nL^{\max})$ (since $\hat{L}^{\max} = \xi L^{\max}$, with $\xi \in [1, n-1]$, as observed
 647 above), where $L^{\max} \leq 2L$ from (2.2).

648 These results seem in agreement with the unconstrained case, where cyclic coordi-
 649 nate selection achieves worse convergence rate than random selection and Gauss-
 650 Southwell-type rules [2, 44], even if practical performances of the algorithms usually
 651 depend on the specific features of the problems.

652 **7.3. Computation of the active-set complexity.** Using all the previous re-
 653 sults, we can now compute the active-set complexity of AC2CD, that is, the maximum
 654 number of iterations required by the algorithm to identify the active set. In particular,
 655 we give an upper bound for \bar{k} appearing in Theorem 6.4 under convexity of f and a
 656 quadratic growth condition, which is now described.

657 We assume that there exists $\mu > 0$ such that

$$658 \quad (7.13) \quad f(x) - f^* \geq \frac{\mu}{2} \|x - x^*\|^2, \quad \forall x \in \mathcal{L}^0,$$

659 where $x^* \in X^*$. Note that (7.13) is automatically satisfied if f is μ -strongly convex
 660 over \mathcal{L}^0 [41]. However, (7.13) is a weaker condition than strong convexity of f over
 661 \mathcal{L}^0 , since there exist convex functions that satisfy (7.13) even if they are non-strongly
 662 convex. This can be seen in the following example, obtained from [39] with proper
 663 adjustments. Note that, in the provided example, f is not even strictly convex, there
 664 is a unique optimal solution x^* (so that $\{x^k\} \rightarrow x^*$) and Assumption 1 is satisfied.

665 **EXAMPLE 1.** Consider the following convex problem:

$$\begin{aligned} \min f(x) &= \frac{1}{2}x_1^2 + \sum_{i=2}^n x_i \\ e^T x &= 0 \\ x_1 &\geq -1 \\ x_i &\geq 0, \quad i = 2, \dots, n, \end{aligned}$$

666

667 with arbitrary dimension $n \geq 3$. Since the smallest eigenvalue of the Hessian matrix
 668 of f is equal to 0, then f is not strongly convex. Actually, f is not even strictly
 669 convex, since $f(\omega x' + (1 - \omega)x'') = \omega f(x') + (1 - \omega)f(x'')$ for all $\omega \in [0, 1]$ and any
 670 distinct feasible points x', x'' such that $x'_1 = x''_1$. We also have that $x^* = 0$ is the
 671 unique optimal solution and $f^* = 0$. We conclude that (7.13) is satisfied with $\mu = 1$,
 672 since $f(x) - f^* = \frac{1}{2}x_1^2 + \sum_{i=2}^n x_i \geq \frac{1}{2} \sum_{i=1}^n x_i^2 = \frac{1}{2} \|x - x^*\|^2$ for all feasible x .

673 **THEOREM 7.5.** The following upper bound holds for \bar{k} appearing in Theorem 6.4
 674 if f is convex over \mathbb{R}^n and satisfies (7.13):

$$675 \quad \bar{k} \leq \left\lfloor \frac{2C}{\mu} \max \left\{ \left(\frac{\tau}{\tau + 1} D^{\max}(x^*) \right)^{-2}, \left(\frac{\zeta(x^*)}{2L + \max \left\{ \frac{1}{A_l}, \frac{L^{\max}}{2(1 - \gamma)} \right\}} \right)^{-2} \right\} \right\rfloor + 1,$$

676 where $C \geq 0$ is the constant of the sublinear convergence rate defined in Theorem 7.4,
 677 $D^{\max}(x^*) > 0$ is the maximum distance from the bounds at x^* , defined as in Proposi-
 678 tion 6.2, $\zeta(x^*) > 0$ is the minimum strict complementarity measure at x^* , defined as
 679 in (6.5), L is the Lipschitz constant of ∇f , $A_l > 0$ is the lower bound on the param-
 680 eter $A^{k,i}$ used to compute the initial stepsize $\Delta^{k,i}$ in the Armijo line search (see (3.6)
 681 and (3.8)), $L^{\max} > 0$ is the maximum among the local Lipschitz constants $L_{i,j}$, defined
 682 as in (2.4), and $\tau \in (0, 1]$ is the parameter used to choose $j(k)$, satisfying (3.3).

683 *Proof.* By the definition of \bar{k} given in Theorem 6.4, it holds that $\bar{k} \geq \max\{k^j, k^z\}$
 684 and (6.6) is satisfied. Recalling the definition of k^z given in Proposition 6.3 and the
 685 fact that (7.1) holds for all $k \geq 0$, we have $k^z = 0$, and then $\bar{k} \geq \max\{k^j, k^z\} = k^j$.
 686 So, from (6.6) and the definition of k^j given in Proposition 6.2, it follows that \bar{k} is

687 the first outer iteration such that

$$688 \quad (7.14a) \quad \|x^k - x^*\|_\infty < \frac{\tau}{\tau+1} D^{\max}(x^*), \quad \forall k \geq \bar{k}.$$

$$689 \quad (7.14b) \quad \|z^{k,i} - x^*\| < \frac{\zeta(x^*)}{2L + \max\left\{\frac{1}{A_i}, \frac{L^{\max}}{2(1-\gamma)}\right\}}, \quad i = 1, \dots, n, \quad \forall k \geq \bar{k}.$$

690
691 By Theorem 7.4 and (7.13), for all $k \geq 1$ we hence have that

$$692 \quad \|x^k - x^*\|_\infty^2 \leq \|x^k - x^*\|^2 \leq \frac{2}{\mu} [f(x^k) - f^*] \leq \frac{2C}{\mu k},$$

$$693 \quad \|z^{k,i} - x^*\|^2 \leq \frac{2}{\mu} [f(z^{k,i}) - f^*] \leq \frac{2}{\mu} [f(x^k) - f^*] \leq \frac{2C}{\mu k}, \quad i = 1, \dots, n,$$

694
695 where, in the last chain of inequalities, we used the fact that $f(z^{k,i+1}) \leq f(z^{k,i}) \leq$
696 $f(x^k)$, $i = 1, \dots, n$. Therefore, (7.14) holds for all k such that $\sqrt{2C/(\mu k)}$ is less than
697 both the right-hand side of (7.14a) and the right-hand side of (7.14b), yielding to the
698 upper bound for \bar{k} given in the assertion. \square

699 We remark that Theorem 7.5 requires convexity and quadratic growth, but it
700 uses the convergence rate result stated in Theorem 7.4, holding for general convex
701 objective functions. As a consequence, we expect the upper bound provided for \bar{k} in
702 Theorem 7.5 to be loose. Improving the convergence rate of the algorithm under the
703 additional quadratic growth condition may hence be a challenging question, since it
704 affects the active-set complexity.

705 **8. Additional results.** So far we have shown that AC2CD identifies $\mathcal{Z}^+(x^*)$
706 in a finite number \bar{k} of outer iterations (provided $\{x^k\} \rightarrow x^*$), also giving an upper
707 bound for \bar{k} when f is convex and satisfies a quadratic growth condition.

708 Now, we want to show that the counterparts of these results hold as well, in the
709 sense that AC2CD is able to identify the complement of $\mathcal{Z}(x^*)$, the so called *non-*
710 *active set*, in a finite number \hat{k} of outer iterations, where an upper bound for \hat{k} can
711 be computed when f is convex and satisfies (7.13). More specifically, still considering
712 a sequence $\{x^k\} \rightarrow x^*$, we want to show that, for all $k > \hat{k}$,

$$713 \quad (8.1) \quad x_h^k \in (l_h, u_h), \quad \forall h \notin \mathcal{Z}(x^*).$$

714 Actually, (8.1) is quite obvious (it follows from the properties of the limit), but ob-
715 taining an upper bound for \hat{k} can be of interest. In particular, if (6.1) and (8.1) hold
716 for $k > \bar{k}$ and $k > \hat{k}$, respectively, for all $k > \max\{\bar{k}, \hat{k}\}$ we have that

$$717 \quad \mathcal{Z}^+(x^*) \subseteq \{i: x_i^k \in \{l_i, u_i\}\} \subseteq \mathcal{Z}(x^*).$$

718 As a consequence, if x^* is non-degenerate, for all $k > \max\{\bar{k}, \hat{k}\}$ it holds

$$719 \quad (8.2) \quad x_h^k \in \{l_h, u_h\} \Leftrightarrow h \in \mathcal{Z}(x^*),$$

720 that is, the active set is exactly identified after $\max\{\bar{k}, \hat{k}\}$ outer iterations.

721 First we show that (8.1) holds for all sufficiently k , without any assumption on
722 the convexity of f , provided the whole sequence $\{x^k\}$ converges.

723 THEOREM 8.1. Let $\{x^k\}$ be a sequence of points produced by AC2CD and assume
 724 that $\lim_{k \rightarrow \infty} x^k = x^*$. Define the minimum non-zero distance from the bounds at x^*
 725 as

$$726 \quad D^{\min}(x^*) = \min_{i \notin \mathcal{Z}(x^*)} D_i(x^*),$$

727 which is well defined and positive by Assumption 1, and let \hat{k} be the first outer iteration
 728 such that

$$729 \quad \|x^k - x^*\|_\infty < D^{\min}(x^*), \quad \forall k > \hat{k}.$$

730 Then, for all $k > \hat{k}$ we have that

$$731 \quad x_h^k \in (l_h, u_h), \quad \forall h \notin \mathcal{Z}(x^*).$$

732 *Proof.* Consider an outer iteration $k > \hat{k}$ and any index $h \notin \mathcal{Z}(x^*)$. We have
 733 $|x_h^k - x_h^*| \leq \|x^k - x^*\|_\infty < D^{\min}(x^*) \leq D_h(x^*)$, implying that

$$734 \quad (8.3) \quad x_h^k - l_h > x_h^* - l_h - D_h(x^*) \quad \text{and} \quad u_h - x_h^k > u_h - x_h^* - D_h(x^*).$$

735 Therefore, we can write

$$736 \quad \begin{aligned} D_h(x^k) &= \min\{x_h^k - l_h, u_h - x_h^k\} \stackrel{(8.3)}{>} \min\{x_h^* - l_h, u_h - x_h^*\} - D_h(x^*) \\ &= D_h(x^*) - D_h(x^*) = 0, \end{aligned}$$

737 that is, $x_h^k \in (l_h, u_h)$. □

738 We finally give an upper bound for \hat{k} under the same assumptions used in The-
 739 orem 7.5. As in the previous section, also here we assume the parameter A_i^k in the
 740 Armijo line search to be computed in order to satisfy, together with (3.8), condi-
 741 tion (7.1) for all $k \geq 0$, as explained in Subsection 7.1.

742 THEOREM 8.2. The following upper bound holds for \hat{k} appearing in Theorem 8.1
 743 if f is convex over \mathbb{R}^n and satisfies (7.13):

$$744 \quad \hat{k} \leq \left\lfloor \frac{2C}{\mu} (D^{\min}(x^*))^{-2} \right\rfloor + 1,$$

745 where $C \geq 0$ is the constant of the sublinear convergence rate defined in Theorem 7.4,
 746 and $D^{\min}(x^*) > 0$ is the minimum non-zero distance from the bounds at x^* , defined
 747 as in Theorem 8.1.

748 *Proof.* Reasoning as in the proof of Theorem 7.5, the desired result follows from
 749 Theorem 8.1 and the fact that $\|x^k - x^*\|_\infty^2 \leq \|x^k - x^*\|^2 \leq \frac{2C}{\mu k}$ for all $k \geq 1$. □

750 The same remarks stated after Theorem 7.5 hold for Theorem 8.2 as well. Namely,
 751 we expect the upper bound provided for \hat{k} to be loose, since it requires convexity and
 752 quadratic growth, but it uses the convergence rate result of Theorem 7.4, holding for
 753 general convex objective functions.

754 Appendix A. Proofs of the technical results of Section 5.

755 *Proof of Lemma 5.2.* For all $x \in \mathcal{F}$ we have $x_j = b - \sum_{i \neq j} x_i$, $j = 1, \dots, n$. So,

$$756 \quad \begin{aligned} v^T(x' - x'') &= \sum_{i \neq j} v_i(x'_i - x''_i) + v_j(x'_j - x''_j) \\ 757 \quad &= \sum_{i \neq j} v_i(x'_i - x''_i) - v_j \left(\sum_{i \neq j} x'_i - \sum_{i \neq j} x''_i \right) = \sum_{i \neq j} (v_i - v_j)(x'_i - x''_i). \quad \square \\ 758 \end{aligned}$$

759 *Proof of Lemma 5.3.* Fix $j \in \{1, \dots, n\}$ and $x', x'' \in \mathcal{F}$. For all $i = 1, \dots, n$ and
 760 $x \in \mathcal{F}$, let $\phi_{i,j,x}$ be the functions appearing in (2.3). Pick any $h \neq j$ and, from known
 761 results on functions with Lipschitz continuous derivatives [41], we can write

$$\begin{aligned} 762 \quad f(x + t(e_h - e_j)) &= \phi_{h,j,x}(t) \leq \phi_{h,j,x}(0) + t\dot{\phi}_{h,j,x}(0) + \frac{L_{h,j}}{2}t^2 \\ 763 \quad &= f(x) + t\nabla f(x)^T(e_h - e_j) + \frac{L_{h,j}}{2}t^2, \quad \forall t \in \mathbb{R}. \\ 764 \end{aligned}$$

765 Using $t = \frac{1}{L_{h,j}}(\nabla_j f(x) - \nabla_h f(x))$, we get

$$766 \quad (\text{A.1}) \quad f(x) - f\left(x + \frac{1}{L_{h,j}}(\nabla_j f(x) - \nabla_h f(x))(e_h - e_j)\right) \geq \frac{1}{2L_{h,j}}(\nabla_h f(x) - \nabla_j f(x))^2.$$

767 Let $\bar{f} = \inf_{x \in \mathbb{R}^n} f(x)$. For all $x \in \mathbb{R}^n$ we can write

$$\begin{aligned} f(x) - \bar{f} &\geq f(x) - f\left(x + \frac{1}{L_{h,j}}(\nabla_j f(x) - \nabla_h f(x))(e_h - e_j)\right) \\ &\geq \frac{1}{2} \max_{i \neq j} \frac{1}{L_{i,j}} (\nabla_i f(x) - \nabla_j f(x))^2 \\ 768 \quad (\text{A.2}) \quad &\stackrel{(*)}{\geq} \frac{1}{2 \sum_{i \neq j} L_{i,j}} \sum_{i=1}^n (\nabla_i f(x) - \nabla_j f(x))^2 \\ &= \frac{1}{2L_j} \sum_{i \neq j} (\nabla_i f(x) - \nabla_j f(x))^2 = \frac{1}{2L_j} \|\nabla f(x) - \nabla_j f(x)e\|_{(j)}^2, \end{aligned}$$

769 where the second inequality follows (A.1), whereas the inequality (*) follows from the
 770 fact that

$$771 \quad \max_{i=1, \dots, r} \frac{a_i}{b_i} \geq \frac{1}{b_1 + \dots + b_r} \sum_{i=1}^n a_i,$$

772 for all $a_1, \dots, a_r \in \mathbb{R}$ and $b_1, \dots, b_r > 0$.

773 Now, define the convex function $\psi_1(x) = f(x) - f(x') - \nabla f(x')^T(x - x')$. Since
 774 $\nabla \psi_1(x) = \nabla f(x) - \nabla f(x')$, for all $x \in \mathcal{F}$, $i \in \{1, \dots, n\}$ and $t, s \in \mathbb{R}$, we can write

$$\begin{aligned} 775 \quad &|\nabla \psi_1(x + t(e_i - e_j))^T(e_i - e_j) - \nabla \psi_1(x + s(e_i - e_j))^T(e_i - e_j)| \\ 776 \quad &= |\nabla f(x + t(e_i - e_j))^T(e_i - e_j) - \nabla f(x + s(e_i - e_j))^T(e_i - e_j)| \leq L_{i,j}|t - s|, \end{aligned}$$

778 where the last inequality follows from the fact that $L_{i,j}$ are local Lipschitz constants
 779 for $\nabla f(x)$. Therefore, $L_{i,j}$ are also local Lipschitz constants for $\nabla \psi_1$. Consequently,
 780 we can use (A.2) with f replaced by ψ_1 . Observing that $\min_{x \in \mathbb{R}^n} \psi_1(x) = 0$, we obtain

$$\begin{aligned} 781 \quad \psi_1(x) &\geq \frac{1}{2L_j} \|\nabla \psi_1(x) - \nabla_j \psi_1(x)e\|_{(j)}^2 \\ 782 \quad &= \frac{1}{2L_j} \|(\nabla f(x) - \nabla_j f(x)e) - (\nabla f(x') - \nabla_j f(x')e)\|_{(j)}^2, \quad \forall x \in \mathbb{R}^n. \\ 783 \end{aligned}$$

784 Using $x = x''$ in the above relation, we get

$$785 \quad (\text{A.3}) \quad \psi_1(x'') \geq \frac{1}{2L_j} \|(\nabla f(x'') - \nabla_j f(x'')e) - (\nabla f(x') - \nabla_j f(x')e)\|_{(j)}^2.$$

786 Defining the function $\psi_2(x) = f(x) - f(x'') - \nabla f(x'')^T(x - x'')$, we can reason as
787 above and we obtain

$$788 \quad (\text{A.4}) \quad \psi_2(x') \geq \frac{1}{2L_j} \left\| (\nabla f(x') - \nabla_j f(x')e) - (\nabla f(x'') - \nabla_j f(x'')e) \right\|_{\langle j \rangle}^2.$$

789 Summing (A.3) and (A.4), we get

$$790 \quad \left\| [\nabla f(x') - \nabla_j f(x')e] - [\nabla f(x'') - \nabla_j f(x'')e] \right\|_{\langle j \rangle}^2 \leq L_j [\nabla f(x') - \nabla f(x'')]^T(x' - x'').$$

791 So, to obtain the desired result we have to show that $[\nabla f(x') - \nabla f(x'')]^T(x' - x'')$ is
792 less than or equal to

$$793 \quad (\text{A.5}) \quad \left\| [\nabla f(x') - \nabla_j f(x')e] - [\nabla f(x'') - \nabla_j f(x'')e] \right\|_{\langle j \rangle} \|x' - x''\|_{\langle j \rangle}.$$

794 This can be achieved by using Lemma 5.2 first with $v = \nabla f(x')$ and then with
795 $v = \nabla f(x'')$, in order to rewrite $[\nabla f(x') - \nabla f(x'')]^T(x' - x'')$ as

$$796 \quad \langle [\nabla f(x') - \nabla_j f(x')e] - [\nabla f(x'') - \nabla_j f(x'')e], x' - x'' \rangle_j.$$

797 Hence, by using inequality (2.8) we obtain that the above quantity is less than or
798 equal to (A.5). \square

799 *Proof of Corollary 5.4.* From (2.9) and the definition of $g^{k,i}$ given in (3.4), for all
800 $v \in \mathbb{R}^n$ we have that

$$801 \quad |\nabla_{p_i^k} f(v) - \nabla_{j(k)} f(v) + g^{k,i}| \leq \left\| [\nabla f(v) - \nabla_{j(k)} f(v)e] - [\nabla f(z^{k,i}) - \nabla_{j(k)} f(z^{k,i})e] \right\|_{\langle j(k) \rangle}.$$

802 Using Lemma 5.3, the desired result is obtained. \square

803 *Proof of Lemma 5.5.* Fix $j \in \{1, \dots, n\}$ and $x', x'' \in \mathcal{F}$. From the mean value
804 theorem and using Lemma 5.2 with $v = \nabla f(x' + t(x'' - x'))$, we have

$$805 \quad \begin{aligned} f(x'') - f(x') &= \int_0^1 \nabla f(x' + t(x'' - x'))^T(x'' - x') dt \\ &= \int_0^1 \langle \nabla f(x' + t(x'' - x')) - \nabla_j f(x' + t(x'' - x'))e, x'' - x' \rangle_j dt. \end{aligned}$$

806 The integrand in the last term of the above chain of equalities can be rewritten as the
807 sum of $\langle \nabla f(x') - \nabla_j f(x')e, x'' - x' \rangle_j$ and

$$808 \quad \langle [\nabla f(x' + t(x'' - x')) - \nabla_j f(x' + t(x'' - x'))e] - [\nabla f(x') - \nabla_j f(x')e], x'' - x' \rangle_j,$$

809 and the latter, by using inequality (2.8) and Lemma 5.3, is less than or equal to
810 $tL_j \|x' - x''\|_{\langle j \rangle}^2$. Therefore,

$$811 \quad \begin{aligned} f(x'') &\leq f(x') + \langle \nabla f(x') - \nabla_j f(x')e, x'' - x' \rangle_j + L_j \|x' - x''\|_{\langle j \rangle}^2 \int_0^1 t dt \\ &= f(x') + \langle \nabla f(x') - \nabla_j f(x')e, x'' - x' \rangle_j + \frac{L_j}{2} \|x' - x''\|_{\langle j \rangle}^2. \end{aligned}$$

812 Using Lemma 5.2 with $v = \nabla f(x')$, the desired result is obtained. \square

813 **Acknowledgments.** The author would like to thank the two anonymous re-
814 viewers for their comments and suggestions.

815

REFERENCES

- 816 [1] A. BECK, *The 2-coordinate descent method for solving double-sided simplex constrained mini-*
817 *mization problems*, J. Optim. Theory Appl., 162 (2014), pp. 892–919.
- 818 [2] A. BECK AND L. TETRUASHVILI, *On the convergence of block coordinate descent type methods*,
819 SIAM J. Optim., 23 (2013), pp. 2037–2060.
- 820 [3] D. P. BERTSEKAS, *On the Goldstein-Levitin-Polyak gradient projection method*, IEEE Trans.
821 Automat. Control, 21 (1976), pp. 174–184.
- 822 [4] E. G. BIRGIN AND J. M. MARTÍNEZ, *Large-scale active-set box-constrained optimization method*
823 *with spectral projected gradients*, Comput. Optim. Appl., 23 (2002), pp. 101–125.
- 824 [5] I. M. BOMZE, F. RINALDI, AND S. R. BULÒ, *First-order Methods for the Impatient: Support*
825 *Identification in Finite Time with Convergent Frank–Wolfe Variants*, SIAM J. Optim., 29
826 (2019), pp. 2211–2226.
- 827 [6] I. M. BOMZE, F. RINALDI, AND D. ZEFFIRO, *Active Set Complexity of the Away-Step Frank-*
828 *Wolfe Algorithm*, SIAM J. Optim., 30 (2020), pp. 2470–2500.
- 829 [7] J. BURKE, *On the identification of active constraints II: The nonconvex case*, SIAM J. Numer.
830 Anal., 27 (1990), pp. 1081–1102.
- 831 [8] J. V. BURKE AND J. J. MORÉ, *On the identification of active constraints*, SIAM J. Numer.
832 Anal., 25 (1988), pp. 1197–1211.
- 833 [9] J. V. BURKE AND J. J. MORÉ, *Exposing constraints*, SIAM J. Optim., 4 (1994), pp. 573–595.
- 834 [10] P. H. CALAMAI AND J. J. MORÉ, *Projected gradient methods for linearly constrained problems*,
835 Math. Program., 39 (1987), pp. 93–116.
- 836 [11] K. L. CLARKSON, *Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm*, ACM
837 Trans. Algorithms, 6 (2010), pp. 1–30.
- 838 [12] A. CRISTOFARI, *An almost cyclic 2-coordinate descent method for singly linearly constrained*
839 *problems*, Comput. Optim. Appl., 73 (2019), pp. 411–452.
- 840 [13] A. CRISTOFARI, M. DE SANTIS, S. LUCIDI, AND F. RINALDI, *A Two-Stage Active-Set Algorithm*
841 *for Bound-Constrained Optimization*, J. Optim. Theory Appl., 172 (2017), pp. 369–401.
- 842 [14] A. CRISTOFARI, M. DE SANTIS, S. LUCIDI, AND F. RINALDI, *An active-set algorithmic framework*
843 *for non-convex optimization problems over the simplex*, Comput. Optim. Appl., 77 (2020),
844 pp. 57–89.
- 845 [15] A. CRISTOFARI, F. RINALDI, AND F. TUDISCO, *Total variation based community detection using*
846 *a nonlinear optimization approach*, SIAM J. Appl. Math., 80 (2020), pp. 1392–1419.
- 847 [16] A. DANILIDIS, C. SAGASTIZÁBAL, AND M. SOLODOV, *Identifying structure of nonsmooth convex*
848 *functions by the bundle technique*, SIAM J. Optim., 20 (2009), pp. 820–840.
- 849 [17] M. DE SANTIS, S. LUCIDI, AND F. RINALDI, *A Fast Active Set Block Coordinate Descent Algo-*
850 *algorithm for ℓ_1 -Regularized Least Squares*, SIAM J. Optim., 26 (2016), pp. 781–809.
- 851 [18] J. C. DUCHI, F. RUAN, ET AL., *Asymptotic optimality in stochastic optimization*, Ann. Statist.,
852 49 (2021), pp. 21–48.
- 853 [19] J. C. DUNN, *On the convergence of projected gradient processes to singular critical points*, J.
854 Optim. Theory Appl., 55 (1987), pp. 203–216.
- 855 [20] F. FACCHINEI, J. JÚDICE, AND J. SOARES, *An active set Newton algorithm for large-scale*
856 *nonlinear programs with box constraints*, SIAM J. Optim., 8 (1998), pp. 158–186.
- 857 [21] F. FACCHINEI AND J.-S. PANG, *Finite-dimensional variational inequalities and complementarity*
858 *problems*, Springer Science & Business Media, 2003.
- 859 [22] E. M. GAFNI AND D. P. BERTSEKAS, *Two-metric projection methods for constrained optimiza-*
860 *tion*, SIAM J. Control Optim., 22 (1984), pp. 936–964.
- 861 [23] W. W. HAGER AND H. ZHANG, *A new active set algorithm for box constrained optimization*,
862 SIAM J. Optim., 17 (2006), pp. 526–557.
- 863 [24] W. HARE, *Identifying active manifolds in regularization problems*, in Fixed-Point Algorithms
864 for Inverse Problems in Science and Engineering, Springer, 2011, pp. 261–271.
- 865 [25] W. L. HARE, *A proximal method for identifying active manifolds*, Comput. Optim. Appl., 43
866 (2009), pp. 295–306.
- 867 [26] W. L. HARE AND A. S. LEWIS, *Identifying active constraints via partial smoothness and prox-*
868 *regularity*, J. Convex Anal., 11 (2004), pp. 251–266.
- 869 [27] C.-J. HSIEH, K.-W. CHANG, C.-J. LIN, S. S. KEERTHI, AND S. SUNDARARAJAN, *A dual coordi-*
870 *nate descent method for large-scale linear SVM*, in Proceedings of the 25th international
871 conference on Machine learning, 2008, pp. 408–415.
- 872 [28] M. JAGGI AND S. LACOSTE-JULIEN, *On the global linear convergence of frank-wolfe optimization*
873 *variants*, Advances in Neural Information Processing Systems, 28 (2015).
- 874 [29] S. LEE AND S. J. WRIGHT, *Manifold identification in dual averaging for regularized stochastic*
875 *online learning*, J. Mach. Learn. Res., 13 (2012), pp. 1705–1744.

- 876 [30] A. S. LEWIS AND S. J. WRIGHT, *Identifying activity*, SIAM J. Optim., 21 (2011), pp. 597–614.
- 877 [31] R. M. LEWIS AND V. TORCZON, *Active set identification for linearly constrained minimization*
- 878 *without explicit derivatives*, SIAM J. Optim., 20 (2010), pp. 1378–1405.
- 879 [32] J. LIANG, J. FADILI, AND G. PEYRÉ, *Activity Identification and Local Linear Convergence of*
- 880 *Forward–Backward-type Methods*, SIAM J. Optim., 27 (2017), pp. 408–437.
- 881 [33] C.-J. LIN, S. LUCIDI, L. PALAGI, A. RISI, AND M. SCIANDRONE, *Decomposition algorithm*
- 882 *model for singly linearly-constrained problems subject to lower and upper bounds*, J. Optim.
- 883 *Theory Appl.*, 141 (2009), pp. 107–126.
- 884 [34] Z.-Q. LUO AND P. TSENG, *On the convergence of the coordinate descent method for convex*
- 885 *differentiable minimization*, J. Optim. Theory Appl., 72 (1992), pp. 7–35.
- 886 [35] Z.-Q. LUO AND P. TSENG, *On the convergence rate of dual ascent methods for linearly con-*
- 887 *strained convex minimization*, Math. Oper. Res., 18 (1993), pp. 846–867.
- 888 [36] R. MIFFLIN AND C. SAGASTIZÁBAL, *Proximal points are on the fast track*, J. Convex Anal., 9
- 889 (2002), pp. 563–580.
- 890 [37] I. NECOARA, *Random coordinate descent algorithms for multi-agent convex optimization over*
- 891 *networks*, IEEE Trans. Automat. Control, 58 (2013), pp. 2001–2012.
- 892 [38] I. NECOARA, Y. NESTEROV, AND F. GLINEUR, *Random block coordinate descent methods for*
- 893 *linearly constrained optimization over networks*, Journal of Optimization Theory and Ap-
- 894 *lications*, 173 (2017), pp. 227–254.
- 895 [39] I. NECOARA, Y. NESTEROV, AND F. GLINEUR, *Linear convergence of first order methods for*
- 896 *non-strongly convex optimization*, Math. Program., 175 (2019), pp. 69–107.
- 897 [40] I. NECOARA AND A. PATRASCU, *A random coordinate descent algorithm for optimization prob-*
- 898 *lems with composite objective function and linear coupled constraints*, Computational Op-
- 899 *timization and Applications*, 57 (2014), pp. 307–337.
- 900 [41] Y. NESTEROV, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer
- 901 *Science & Business Media*, 2013.
- 902 [42] J. NUTINI, I. LARADJI, AND M. SCHMIDT, *Let’s Make Block Coordinate Descent Go Fast:*
- 903 *Faster Greedy Rules, Message-Passing, Active-Set Complexity, and Superlinear Conver-*
- 904 *gence*, preprint, <https://arxiv.org/abs/1712.08859> (2017).
- 905 [43] J. NUTINI, M. SCHMIDT, AND W. HARE, *“Active-set complexity” of proximal gradient: How*
- 906 *long does it take to find the sparsity pattern?*, Optim. Lett., 13 (2019), pp. 645–655.
- 907 [44] J. NUTINI, M. SCHMIDT, I. LARADJI, M. FRIEDLANDER, AND H. KOEPKE, *Coordinate descent*
- 908 *converges faster with the gauss-southwell rule than random selection*, in International Confe-
- 909 *rence on Machine Learning*, PMLR, 2015, pp. 1632–1641.
- 910 [45] J. M. ORTEGA AND W. C. RHEINBOLDT, *Iterative solution of nonlinear equations in several*
- 911 *variables*, vol. 30, Siam, 1970.
- 912 [46] A. PATRASCU AND I. NECOARA, *Efficient random coordinate descent algorithms for large-scale*
- 913 *structured nonconvex optimization*, J. Global Optim., 61 (2015), pp. 19–46.
- 914 [47] C. POON, J. LIANG, AND C. SCHOENLIEB, *Local convergence properties of SAGA/Prox-SVRG*
- 915 *and acceleration*, in Proc. Mach. Learn. Res. (PMLR), 2018, pp. 4124–4132.
- 916 [48] J. SHE AND M. SCHMIDT, *Linear convergence and support vector identification of sequential*
- 917 *minimal optimization*, in 10th NIPS Workshop on Optimization for Machine Learning,
- 918 vol. 5, 2017.
- 919 [49] Y. SUN, H. JEONG, J. NUTINI, AND M. SCHMIDT, *Are we there yet? manifold identification*
- 920 *of gradient-related proximal methods*, in The 22nd International Conference on Artificial
- 921 *Intelligence and Statistics*, 2019, pp. 1110–1119.
- 922 [50] S. J. WRIGHT, *Identifiable surfaces in constrained optimization*, SIAM J. Control Optim., 31
- 923 (1993), pp. 1063–1079.
- 924 [51] S. J. WRIGHT, *Accelerated block-coordinate relaxation for regularized optimization*, SIAM J.
- 925 *Optim.*, 22 (2012), pp. 159–186.