A PROXIMITY CONTROL ALGORITHM TO MINIMIZE NONSMOOTH AND NONCONVEX SEMI-INFINITE MAXIMUM EIGENVALUE FUNCTIONS

DOMINIKUS NOLL*, OLIVIER PROT*, AND PIERRE APKARIAN*†

Abstract. Proximity control is a well-known mechanism in bundle method for nonsmooth optimization. Here we show that it can be used to optimize a large class of nonconvex and nonsmooth functions with additional structure. This includes for instance nonconvex maximum eigenvalue functions, and also infinite suprema of such functions.

Key words. Nonsmooth calculus, nonsmooth optimization, Clarke subdifferential, spectral bundle method, maximum eigenvalue function, semi-infinite problem, H_∞-norm.

1. Introduction. Proximity control for bundle methods has been known for a long time, but its use is too often restricted to convex optimization, where its full strength cannot be gauged. As we shall demonstrate, as soon as the management of the proximity control parameter follows the lines of a trust region strategy, many nonconvex and nonsmooth locally Lipschitz functions can be optimized. In contrast, in the convex case, the proximity control parameter can usually be frozen, which suggests that under convexity the full picture is not seen, and something of the essence is missing to understand this mechanism. The method we discuss here will be developed in the context of a specific application, because that is where the motivation of our work arises from, but we will indicate in which way the method can be generalized to much larger classes of functions.

The application we have in mind is optimizing the H_∞-norm, which is structurally of the form

$$f(x) = \sup_{\omega \in [0,\infty]} \lambda_1 \left(F(x,\omega) \right), \tag{1}$$

where $F : \mathbb{R}^n \times [0,\infty] \to \mathbb{S}^m$ is an operator with values in the space $\mathbb{S}^m$ of $m \times m$ symmetric or Hermitian matrices, equipped with the scalar product $X \bullet Y = \mathrm{Tr}(XY)$, and where λ_1 denotes the maximum eigenvalue function on $\mathbb{S}^m$. We assume that F is jointly continuous in the variable (x,ω) and of class C^2 in the variable x, so that $F''(x,\omega)$ is still jointly continuous. Here derivatives always refer to the variable x. Our exposition will show how these hypotheses can easily be relaxed. The program we wish to solve is

$$\min_{x \in \mathbb{R}^n} f(x) \tag{2}$$

where f has the form (1).

The approach presented here was originally developed in the context of eigenvalue optimization, and [8] gives an overview of the history. The bases for the present extension to the semi-infinite case were laid in [4, 1, 56, 11, 5, 6]. Our method is inspired by Helmberg and Rendl's spectral bundle method [25], where large semidefinite programs arising as relaxations of quadratic integer programming problems are developed. Helmberg and Rendl optimize a convex eigenvalue function of the form

*UNIVERSITÉ PAUL SABATIER, INSTITUT DE MATHÉMATIQUES, 31062 TOULOUSE, FRANCE

†ONERA-CERT, AV. EDOUARD BÉLIN, 31055 TOULOUSE, FRANCE

$\lambda_1(A(x))$, where $A : \mathbb{R}^n \to \mathbb{S}^m$ is affine. This method has also antecedents in classical bundling, like Lemaréchal [35, 36, 37, 39] or Kiwiel [32, 33, 31]. Extensions of the convex case to include bound constraints are given in [23].

Optimization of the H_∞-norm is an important application in feedback control synthesis, which has been pioneered by E. Polak and co-workers. See for instance [42, 43, 52] and the references given there. Our own approach to optimizing the H_∞ norm is developed in [4, 1, 5].

The structure of the paper is as follows. After some preparation in sections 2 and 3, the core of the algorithm is explained in section 5. The algorithm is presented in section 6. Convergence proofs for the inner and outer loop follow in sections 7 and 8.

2. Preparation. Observe that our objective function has the form

$$f(x) = \max_{\omega \in [0,\infty]} f(x,\omega), \tag{3}$$

where each $f(x,\omega) = \lambda_1 \left(F(x,\omega)\right)$ is a composite maximum eigenvalue function. Recall that the maximum eigenvalue function $\lambda_1 : \mathbb{S}^m \to \mathbb{R}$ is the support function of the compact convex set

$$\mathcal{C} = \{Z \in \mathbb{S}^m : Z \succeq 0, \mathrm{Tr}(Z) = 1\},$$

where $\succeq 0$ means positive semidefinite. In other words,

$$f(x) = \max_{\omega \in [0,\infty]} \max_{Z \in \mathcal{C}} Z \bullet F(x,\omega). \tag{4}$$

Due to compactness of $\mathcal{C}$ and $[0, \infty]$, the suprema in (4), are attained. This suggests introducing an approximation of f in a neighbourhood of x, which is

$$\begin{aligned} \phi(y,x) &= \max_{\omega \in [0,\infty]} \lambda_1 \left(F(x,\omega) + F'(x,\omega)(y-x)\right) \\ &= \max_{\omega \in [0,\infty]} \max_{Z \in \mathcal{C}} Z \bullet \left(F(x,\omega) + F'(x,\omega)(y-x)\right) \end{aligned} \tag{5}$$

where the derivative $F'(x,\omega)$ refers to the variable x. As (5) uses a Taylor expansion of the operator F in a neighbourhood of x, we expect $\phi(y,x)$ to be a good model of f for y near x. This is confirmed by the following

LEMMA 1. *Let $B \subset \mathbb{R}^n$ be a bounded set. Then there exists a constant $L > 0$ such that*

$$|f(y) - \phi(y,x)| \leq L\|y - x\|^2$$

for all $x, y \in B$.

Proof. By Weil's theorem we have

$$\lambda_m(E) \leq \lambda_1(A + E) - \lambda_1(A) \leq \lambda_1(E)$$

for all matrices $A, E \in \mathbb{S}^m$. We apply this with $A = F(y,\omega)$ and $A + E = F(x,\omega) + F'(x,\omega)(y - x)$. Now observe that by hypothesis on F there exists $L > 0$ such that

$$\sup_{z \in B} \sup_{\omega \in [0,\infty]} \|F''(z,\omega)\| \leq L.$$

This proves $E = \mathcal{O}(\|y - x\|^2)$, uniformly over $x, y \in B$ and uniformly over $\omega \in [0, \infty]$. □

The following is a specific property of the H_∞-norm, which can be exploited algorithmically. A proof can be found in [13] or [11].

LEMMA 2. *The set* $\Omega(x) = \{\omega \in [0, \infty] : f(x) = f(x, \omega)\}$ *is either finite, or* $\Omega(x) = [0, \infty]$. *We call* $\Omega(x)$ *the set of active frequencies.*

For later use let us mention a different way to represent the convex model $\phi(y, x)$. We introduce the notations

$$\alpha(\omega, Z) = Z \bullet F(x, \omega) \in \mathbb{R}, \qquad g(\omega, Z) = F'(x, \omega)^\star Z \in \mathbb{R}^n.$$

and we let

$$\mathcal{G} = \mathrm{co}\left\{(\alpha(\omega, Z), g(\omega, Z)) : \omega \in [0, \infty], Z \in \mathcal{C}\right\},$$

where $\mathrm{co}(X)$ is the convex hull of X. Then we have the following equivalent representation of the model:

$$\phi(y, x) = \max\{\alpha + g^\top (y - x) : (\alpha, g) \in \mathcal{G}\}. \tag{6}$$

3. Tangent program. Suppose x is the current iterate of our algorithm to be designed. In order to generate trial steps away from x, we will recursively construct approximations $\phi_k(y, x)$ of $\phi(y, x)$ of increasing quality. Using the form (6) we will choose suitable subsets $\mathcal{G}_k$ of the set $\mathcal{G}$ and define

$$\phi_k(y, x) = \max\{\alpha + g^\top (y - x) : (\alpha, g) \in \mathcal{G}_k\}. \tag{7}$$

Clearly $\phi_k \le \phi$, and a suitable strategy will assure that the ϕ_k get closer to the model ϕ as k increases. Once the model $\mathcal{G}_k$ is formed, a new trial step y^{k+1} is generated by solving the tangent program

$$\min_{y \in \mathbb{R}^n} \phi_k(y, x) + \tfrac{\delta_k}{2} \|y - x\|^2, \tag{8}$$

where $\delta_k > 0$ is the proximity control parameter, which will be adjusted anew at each step k. Here we make the implicit assumption that solving (8) is much easier than solving the original problem.

Suppose the solution of (8) is y^{k+1}. Following standard terminology in nonsmooth optimization, y^{k+1} will be called a serious step if it is accepted to become the new iterate x^+. On the other hand, if y^{k+1} is not satisfactory and has to be rejected, it is called a null step. In that case, a new model $\mathcal{G}_{k+1}$ is built, using information from the previous $\mathcal{G}_k$, and integrating information provided by y^{k+1}. The proximity parameter is updated, $\delta_k \to \delta_{k+1}$, and the tangent program is solved again. In other words, the construction of the $\mathcal{G}_k$ in (7) is recursive.

In order to guarantee convergence of our method, we have isolated three basic properties of the sets $\mathcal{G}_k$. The most basic one is certainly that $\phi_k(x, x) = \phi(x, x) = f(x)$, and this is covered by the following:

LEMMA 3. *Let* $\omega_0 \in \Omega(x)$ *be any of the active frequencies at* x. *Choose a normalized eigenvector* e_0 *associated with the maximum eigenvalue* $f(x) = \lambda_1(F(x, \omega_0))$ *of* $F(x, \omega_0)$, *and let* $Z_0 := e_0 e_0^\top \in \mathcal{C}$. *If* $(\alpha(\omega_0, Z_0), g(\omega_0, Z_0)) \in \mathcal{G}_k$, *then* $\phi_k(x, x) = \phi(x, x) = f(x)$. $\square$

A second more sophisticated property of our model $\phi_k(\cdot, x)$ is that it is improved at each step by adding suitable affine support functions of $\phi(\cdot, x)$, referred to as cutting planes. Suppose a trial step y^{k+1} away from x is computed via (8), based on the current model $\phi_k(\cdot, x)$ with approximation $\mathcal{G}_k$ and proximity control parameter δ_k. If y^{k+1} fails because the progress in the function value is not satisfactory (null step), we add an affine support function of $\phi(\cdot, x)$ to the next model $\phi_{k+1}(\cdot, x)$. This will assure that the bad step y^{k+1} will be cut away at the next step $k+1$, hopefully paving the way for something better to come. What we have in mind is made precise by the following:

LEMMA 4. *Let $\omega_{k+1} \in [0, \infty]$ and $Z_{k+1} \in \mathcal{C}$ be where the maximum (5) for the solution y^{k+1} of (8) is attained, that is,*

$$\phi(y^{k+1}, x) = Z_{k+1} \bullet \left(F(x, \omega_{k+1}) + F'(x, \omega_{k+1})(y^{k+1} - x)\right).$$

If $(\alpha(\omega_{k+1}, Z_{k+1}), g(\omega_{k+1}, Z_{k+1})) \in \mathcal{G}_{k+1}$, then we have $\phi_{k+1}(y^{k+1}, x) = \phi(y^{k+1}, x)$.

We need yet another support function to improve the model, and this is usually called the aggregation element. The idea is as follows. As we keep updating our approximation and $\mathcal{G}_k$, we expect our model $\phi_k(\cdot, x)$ to get closer to f. The easiest way to assure this would seem to let the sequence increase: $\mathcal{G}_k \subset \mathcal{G}_{k+1}$, so that previous attempts (null steps) are perfectly memorized. However, this would quickly lead to overload. To avoid this, we drive ϕ_k toward ϕ in a more sophisticated way by a clever use of the information obtained from the null steps. As we have seen, adding a cutting plane avoids the last unsuccessful step y^{k+1}. This could be considered a reality check, where ϕ_k is matched with ϕ. What is further needed is relating ϕ_{k+1} to its past, ϕ_k, and this is what aggregation is about.

According to the definition of y^{k+1} as minimum of the tangent program (8) we have $0 \in \partial\phi_k(y^{k+1}, x) + \delta_k(y^{k+1} - x)$. The way ϕ_k is built (7) shows that this may be written as

$$0 = \sum_{i=1}^{r} \tau_i^* g_i^* + \delta_k(y^{k+1} - x) \tag{9}$$

for certain $\tau_i^* \geq 0$ summing up to 1, and $(\alpha_i^*, g_i^*) \in \mathcal{G}_k$. We let

$$\alpha^* = \sum_{i=1}^{r} \tau_i^* \alpha_i^*, \qquad g^* = \sum_{i=1}^{r} \tau_i^* g_i^*, \tag{10}$$

and keep $(\alpha^*, g^*) \in \mathcal{G}_{k+1}$. Notice that this pair belong indeed to $\mathcal{G}$ by convexity, and because $\mathcal{G}_k \subset \mathcal{G}$.

Altogether, we have now isolated three properties, which our approximations $\mathcal{G}_k$ have to satisfy:

- $(G1)$ $\mathcal{G}_k$ contains at least one pair $(\alpha(\omega_0, Z_0), g(\omega_0, Z_0))$, where $\omega_0 \in \Omega(x)$ is an active frequency, $Z_0 = e_0 e_0^\top$ for a normalized eigenvector e_0 of $F(x, \omega_0)$ associated with $\lambda_1(F(x, \omega_0))$.
- $(G2)$ For every null step y^{k+1}, $\mathcal{G}_{k+1}$ contains a pair $(\alpha(\omega_{k+1}, Z_{k+1}), g(\omega_{k+1}, Z_{k+1}))$, where ω_{k+1}, Z_{k+1} satisfy $\phi(y^{k+1}, x) = Z_{k+1} \bullet [F(x, \omega_{k+1}) + F'(x, \omega_{k+1})(y^{k+1} - x)]$.
- $(G3)$ If $\delta_k(x - y^{k+1}) \in \partial\phi_k(y^{k+1}, x)$ for a null step y^{k+1}, then $\mathcal{G}_{k+1}$ contains the aggregate pair (α^*, g^*) satisfying (9) and (10).

As we shall see, these properties guarantee a weak form of convergence of our method. Practical considerations, however, require richer sets $\mathcal{G}_k$ which in general are no longer finitely generated. The way these are built is explained in the next section. To conclude, we state the consequences of the three axioms in the following

LEMMA 5. *Axioms $(G1)$ - $(G3)$ guarantee that $\phi_k(x, x) = \phi(x, x) = f(x)$, that $\phi_{k+1}(y^{k+1}, x) = \phi(y^{k+1}, x)$, and that $\phi_{k+1}(y^{k+1}, x) \geq \phi_k(y^{k+1}, x)$.*

4. Solving the tangent program. Our numerical experience shows that it is useful to generate approximations $\mathcal{G}_k$ larger than what is required by the minimal axioms $(G1)$ - $(G3)$. More precisely, we will keep the procedures in $(G2)$ and $(G3)$, but improve on $(G1)$.

Consider the case where the set $\Omega(x)$ of active frequencies is finite. We let Ω_k be a finite extension of $\Omega(x)$, enriched along the lines discussed in [4]. For every $\omega \in \Omega_k$, we allow all sets $Z_\omega \in \mathcal{C}$ of the form

$$Z_\omega = Q_\omega Y_\omega Q_\omega^\top, \quad Y_\omega \succeq 0,\ \mathrm{Tr}(Y_\omega) = 1, \tag{11}$$

where the columns of Q_ω are an orthonormal basis of some invariant subspace of $F(x,\omega)$, containing the eigenspace associated with the maximum eigenvalue. This assures axiom $(G1)$, because $\omega_0 \in \Omega_k$ at all times, and because e_0 belongs to the span of the columns of Q_{ω_0}. Similarly, to force $(G2)$, for every null step y^{k+1} we simply have to keep $\omega_{k+1} \in \Omega_{k+1}$ and let the normalized eigenvector e_{k+1} of $F(x,\omega_{k+1}) + F'(x,\omega_{k+1})(y^{k+1} - x)$ associated with λ_1 be in the span of the columns of $Q_{\omega_{k+1}}$. Then

$$\mathcal{G}_k = \{(\alpha(\omega, Z_\omega), g(\omega, Z_\omega)) : \omega \in \Omega_k, Y_\omega \succeq 0, \mathrm{Tr}(Y_\omega) = 1\} \cup \{(\alpha^*, g^*)\}, \tag{12}$$

where (α^*, g^*) is the aggregate from the previous sweep $k-1$. Notice that $\mathrm{co}(\mathcal{G}_k) \not\subset \mathrm{co}(\mathcal{G}_{k+1})$ in general, because the active frequencies change at each step.

Let us now pass to the more practical aspect on how setting up and solving the tangent program (8) at each step. Writing the tangent program in the form

$$\min_{y\in\mathbb{R}^n} \max_{(\alpha,g)\in\mathrm{co}(\mathcal{G}_k)} \alpha + g^\top(y - x) + \frac{\delta_k}{2}\|y - x\|^2$$

we can use Fenchel duality to swap the min and max operators. The then inner infimum over y is unconstrained and can be computed explicitely, which leads to $y = x - \delta_k^{-1} g$. Substituting this back gives the following form of the dual program

$$\max_{(\alpha,g)\in\mathrm{co}(\mathcal{G}_k)} \alpha - \tfrac{1}{2\delta_k}\|g\|^2.$$

This abstract program takes the following more concrete form if we use the sets $\mathcal{G}_k$ in (12):

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{\omega\in\Omega_k} Y_\omega \bullet Q_\omega^\top F(x,\omega) Q_\omega + \tau\alpha^* - \frac{1}{2\delta_k}\left\| \sum_{\omega\in\Omega_k} F'(x,\omega)^\star \left[Q_\omega Y_\omega Q_\omega^\top\right] + \tau g^* \right\|^2 \\ \text{subject to} & \tau \geq 0, Y_\omega \succeq 0 \\ & \displaystyle\tau + \sum_{\omega\in\Omega_k} \mathrm{Tr}(Y_\omega) = 1 \end{array}$$

The reader will recognize this as a semidefinite program. The return formula takes the explicit form

$$y^{k+1} = x - \frac{1}{\delta_k}\left(\sum_{\omega\in\Omega_k} F'(x,\omega)^\star \left[Q_\omega Y_\omega^* Q_\omega^\top)\right] + \tau^* g^* \right), \tag{13}$$

where (Y^*, τ^*) is the dual optimal solution.

Finally, if we assume that the multiplicity of each maximum eigenvalue is 1, we may further simplify the dual program. This is most often the case in practice. Indeed, in this case the matrices $Z_\omega = e_\omega y_\omega e_\omega^\top$ are of rank 1, so in particular $y_\omega = 1$ is scalar. In other words, we have a finite set of $\alpha_\omega = e_\omega^\top F(x,\omega)e_\omega$ and $g_\omega = F'(x,\omega)^\star e_\omega e_\omega^\top$, $\omega \in \Omega_k$, to which we add the aggregate element (α^*, g^*), and where ω_k required for the last cutting plane is included in Ω_k to assure $(G2)$. Arranging this finite set into a sequence $r = 1, \dots, R_k$, we can write ϕ_k as

$$\phi_k(y,x) = \max_{r=1,\dots,R_k} \alpha_r + g_r^\top (y-x),$$

where $R_k = |\Omega_k| + 1$.

Solving the tangent program at stage k can now be obtain by convex duality. We have the primal form of (8):

$$\min_{y\in\mathbb{R}^n} \max_{r=1,\dots,R_k} \alpha_r + g_r^\top (y-x) + \tfrac{\delta_k}{2}\|y-x\|^2.$$

Standard convex duality shows that the concave dual of this is

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{r=1}^{R_k} \tau_r\alpha_r - \frac{1}{2\delta_k}\left\|\sum_{r=1}^{R_k}\tau_r g_r\right\|^2 \\ \text{subject to} & \displaystyle\sum_{r=1}^{R_k} \tau_r = 1 \\ & 0 \le \tau_r \le 1, r = 1,\dots,R_k \end{array}$$

with unknown variable τ. The return formula to recover the solution of the primal from the solution of the dual is

$$y^{k+1} = x - \frac{1}{\delta_k}\sum_{r=1}^{R_k} \tau_r^* g_r,$$

where τ^* is the optimal solution of the dual.

5. Management of the proximity parameter. At the core of our method is the management of the proximity control parameter δ_k in (8). In order to decide whether the solution y^{k+1} of (8) can be accepted as the new iterate x^+, we compute the control parameter

$$\rho_k = \frac{f(x) - f(y^{k+1})}{f(x) - \phi_k(y^{k+1},x)},$$

which relates our current model $\phi_k(\cdot,x)$ to the truth f. If $\phi_k(\cdot,x)$ is a good model of f, we expect $\rho_k \approx 1$. But we accept y^{k+1} already when $\rho_k \ge \gamma$, (serious step), where the reader might for instance imagine $\gamma = .25$. We say that the agreement between f and ϕ_k is good when $\rho_k \ge \Gamma$, where $\Gamma = .75$ makes sense, and we call it bad when $\rho_k < \gamma$. So we accept steps which are *not bad*. Notice that bad includes in particular those cases where $\rho_k < 0$. As the denominator in ρ_k is always > 0, $\rho_k < 0$ corresponds to those cases where y^{k+1} is not even a descent step for f.

The question is what we should do when y^{k+1} is bad (null step). Here we compute a second control quotient

$$\widetilde{\rho}_k = \frac{f(x) - \phi(y^{k+1},x)}{f(x) - \phi_k(y^{k+1},x)}$$

which compares the models ϕ and ϕ_k. Introduce a similar parameter $\widetilde{\gamma} \in (0,1)$, where $\gamma < \widetilde{\gamma}$, but typically only slightly. We say that agreement between ϕ and ϕ_k is bad if $\widetilde{\rho}_k < \widetilde{\gamma}$, and not bad otherwise. Our decision is now as follows. If $\rho_k < \gamma$ and also $\widetilde{\rho}_k < \widetilde{\gamma}$, then we keep the proximity control parameter unchanged and rely on cutting planes and aggregation, being reluctant to increase δ_k prematurely. Instead we rely on driving ϕ_k closer to ϕ, hoping that this will also bring it closer to f. On the other hand, when $\rho_k < \gamma$, but $\widetilde{\rho}_k \geq \widetilde{\gamma}$, then we are in a more delicate situation where ϕ_k is already reasonably close to ϕ, yet our trial steps do not work because ϕ itself is far from the truth f. Here it will not do to solely keep driving ϕ_k closer to ϕ. We also need to bring $\phi(\cdot, x)$ closer to f. This could only be achieved by tightening proximity control, that is, by increasing δ_k. This is what is done in step 7 of the algorithm. Notice however that even here we continue driving ϕ_k toward ϕ via cutting planes and aggregation, so this process is never stopped.

Finally, if a serious step is accepted with $\rho_k > \Gamma$, we can take confidence in our model, and this is where we relax proximity control by reducing δ_k for the next sweep. This is arranged in step 4 of the algorithm. It may therefore happen that by a succession of such successful steps δ_k approaches 0. This in indeed the ideal case, which in a trust region context corresponds to the case where the trust region constraint becomes inactive.

Even though it is well-known, it is useful to compare the proximity control model (8) to the trust region approach

$$\begin{array}{ll} \text{minimize} & \phi_k(y,x) \\ \text{subject to} & \|y-x\| \leq t_k \end{array} \tag{14}$$

where t_k is the trust region radius. Indeed, as is well-known (see e.g. [28, II, Prop. 2.2.3, p. 291]) solutions of (8) and (14) are in one-to-one correspondence in the sense that if y^{k+1} solves (14) such that the constraint is active with Lagrange multiplier $\lambda_k > 0$, then y^{k+1} solves (8) with $\delta_k = \lambda_k$. Conversely, if y^{k+1} solves (8) with proximity parameter δ_k, then it solves (14) with $t_k = \|y^{k+1} - x\|$. It is now clear that increasing δ_k corresponds to decreasing t_k, and conversely.

Remark. In [12] and [41] the authors discuss tangent program (8) with $\delta_k\|x\|^2$ replaced by more general quadratic forms $x^\top Q_k x$. While this is covered by our convergence theory, the interesting point is that in the cited work the quadratic term is rather interpreted as a substitute of the Hessian of the objective f. In consequence, BFGS updates are proposed for $Q_k \succ 0$. The idea behind this strategy is to speed up convergence in the neighbourhood of the optimum. Since the authors consider convex objectives only, they do not consider the trust region point of view, which is central in our work. The question is then, are these two approaches contradictory? It turns out that we can combine the two ideas by solving a tangent program of the form

$$\min_{y \in \mathbb{R}^n} \phi_k(y, x^\ell) + \frac{1}{2}(y - x^\ell)^\top Q_\ell (y - x^\ell) + \frac{\delta_k}{2}\|y - x^\ell\|^2$$

where x^ℓ are the iterates of the outer loop, y^{k+1} those of the inner loop. We manage δ_k as before, while Q_ℓ is fixed in the inner loop, and updated after every serious step. The additional quadratic term can be integrated in the proofs of section 7. For the convergence of the outer loop in section 8, additional hypotheses implying boundedness of the sequence Q^ℓ and $(Q^\ell)^{-1}$ will be required, just as in the two cited references.

6. The algorithm. In this section we present our algorithm.

Proximity control algorithm for $\min_{x\in\mathbb{R}^n}\max_{\omega\in[0,\infty]} f(x,\omega)$

Parameters $0<\gamma<\widetilde{\gamma}<\Gamma<1$.

0. **Initialize outer loop**. Choose initial x such that $f(x)<\infty$.
1. **Outer loop**. Stop at the current x if $0\in\partial f(x)$. Otherwise compute $\Omega(x)$ and continue with inner loop.
2. **Initialize inner loop**. Choose initial approximation $\mathcal{G}_1$, which contains at least $(\alpha(\omega_0,Z_0),g(\omega_0,Z_0))$, where $\omega_0\in\Omega(x)$ and e_0 is normalized eigenvector associated with $\lambda_1(F(x,\omega_0))$. Possibly enrich $\mathcal{G}_1$ as in (12) via finite extension $\Omega_1\supset\Omega(x)$. Initialize $\delta_1>0$. If old memory element for δ is available, use it to initialize δ_1. Put inner loop counter $k=1$.
3. **Trial step**. At inner loop counter k for given $\mathcal{G}_k$ and proximity parameter δ_k, solve tangent program
$$\min_{y\in\mathbb{R}^n}\phi_k(y,x)+\frac{\delta_k}{2}\|y-x\|^2.$$
The solution is y^{k+1}.
4. **Test of progress**. Check whether
$$\rho_k=\frac{f(x)-f(y^{k+1})}{f(x)-\phi_k(y^{k+1},x)}\geq\gamma.$$
If this is the case, accept trial step y^{k+1} as the new iterate x^+ (serious step). Compute new memory element:
$$\delta^+=\begin{cases}\frac{\delta_k}{2} & \text{if } \rho_k>\Gamma\\ \delta_k & \text{otherwise}\end{cases}$$
and go back to step 1. If $\rho_k<\gamma$ continue with step 5 (null step).
5. **Cutting plane**. Select a frequency ω_{k+1} where $\phi(y^{k+1},x)$ is active and pick a normalized eigenvector e_{k+1} associated with the maximum eigenvalue of $F(x,\omega_{k+1})+F'(x,\omega_{k+1})(y^{k+1}-x)$. Assure $\Omega_{k+1}\supset\Omega(x)\cup\{\omega_0,\omega_{k+1}\}$ and that e_{k+1} is among the columns of $Q_{\omega_{k+1}}$, e_0 among the columns of Q_{ω_0}. Possibly enrich $\mathcal{G}_{k+1}$ as in (12) by adding more frequencies to Ω_{k+1}.
6. **Aggregation**. Compute aggregate pair (α^*,g^*) via (9), (10) based on y^{k+1}, and keep $(\alpha^*,g^*)\in\mathcal{G}_{k+1}$.
7. **Proximity control**. Compute control parameter
$$\widetilde{\rho}_k=\frac{f(x)-\phi(y^{k+1},x)}{f(x)-\phi_k(y^{k+1},x)}.$$
Update proximity parameter δ_k as
$$\delta_{k+1}=\begin{cases}\delta_k, & \text{if } \widetilde{\rho}_k<\widetilde{\gamma}\\ 2\delta_k & \text{if } \widetilde{\rho}_k\geq\widetilde{\gamma}\end{cases}$$
Increase inner loop counter k and go back to step 3.

7. Finiteness of inner loop. We have to show that the inner loop terminates after a finite number of updates k with a new iterate $y^{k+1}=x^+$. This will be proved in the next two Lemmas.

LEMMA 6. *Suppose the inner loop creates an infinite sequence y^{k+1} of null steps with $\rho_k<\gamma$. Then there must be an instant k_0 such that the control parameter $\widetilde{\rho}_k$ satisfies $\widetilde{\rho}_k<\widetilde{\gamma}$ for all $k\geq k_0$.*

Proof. Indeed, by assumption none of the trial steps y^{k+1} passes the acceptance test in step 4, so $\rho_k < \gamma$ at all times k. Suppose now that $\widetilde{\rho}_k \geq \widetilde{\gamma}$ for an infinity of times k. Then according to step 7 the proximity parameter δ_k is increased infinitely often, meaning $\delta_k \to \infty$.

Using the fact that y^{k+1} is the optimal solution of the tangent program gives $0 \in \partial\phi_k(y^{k+1}, x) + \delta_k(y^{k+1} - x)$. Using convexity of ϕ_k, we deduce that

$$-\delta_k(y^{k+1} - x)^\top(x - y^{k+1}) \leq \phi_k(x, x) - \phi_k(y^{k+1}, x)$$

Using $\phi_k(x, x) = f(x)$, assured by keeping $\omega_0 \in \Omega_k$ and $Z_0 \in \mathcal{C}_k$ at all times (Lemma 3), we deduce

$$\frac{\delta_k\|y^{k+1} - x\|^2}{f(x) - \phi_k(y^{k+1}, x)} \leq 1. \tag{15}$$

Now we expand

$$\begin{aligned}
\widetilde{\rho}_k &= \rho_k + \frac{f(y^{k+1}) - \phi(y^{k+1}, x)}{f(x) - \phi_k(y^{k+1}, x)} \\
&\leq \rho_k + \frac{L\|y^{k+1} - x\|^2}{f(x) - \phi_k(y^{k+1}, x)} \qquad \text{(using Lemma 1)} \\
&\leq \rho_k + \frac{L}{\delta_k} \qquad \text{(using (15))}
\end{aligned}$$

Since $L/\delta_k \to 0$, we have $\limsup \widetilde{\rho}_k \leq \limsup \rho_k \leq \gamma < \widetilde{\gamma}$, which contradicts $\widetilde{\rho}_k \geq \widetilde{\gamma}$ for infinitely many k. □

So far we know that if the inner loop turns forever, this implies that $\rho_k < \gamma$ and $\widetilde{\rho}_k < \widetilde{\gamma}$ from some counter k_0 onwards. We show that this cannot happen, by proving the following

LEMMA 7. *Suppose $\rho_k < \gamma$ and $\widetilde{\rho}_k < \widetilde{\gamma}$ for all $k \geq k_0$. Then $0 \in \partial f(x)$.*

Proof. 1) Step 7 of the algorithm tells us that we are in the case where the proximity parameter is no longer increased, and remains therefore constant. Let us say $\delta := \delta_k$ for all $k \geq k_0$.

2) For later use, let us introduce the function

$$\psi_k(y, x) = \phi_k(y, x) + \tfrac{\delta}{2}\|y - x\|^2.$$

As we have seen already, the necessary optimality condition for the tangent program imply

$$\delta\|y^{k+1} - x\|^2 \leq f(x) - \phi_k(y^{k+1}, x).$$

Now remember that in step 6 of the algorithm, and according to axiom $(G3)$, we have kept the aggregate pair $(\alpha^*, g^*) \in \mathcal{G}_{k+1}$. By its definition (9), (10) we have

$$\phi_k(y^{k+1}, x) = \alpha^* + g^{*\top}(y^{k+1} - x).$$

Defining a new function

$$\psi_k^*(y, x) := \alpha^* + g^{*\top}(y - x) + \tfrac{\delta}{2}\|y - x\|^2$$

we therefore have

$$\psi_k^*(y^{k+1}, x) = \psi_k(y^{k+1}, x) \text{ and } \psi_k^*(y, x) \leq \psi_{k+1}(y, x), \tag{16}$$

the latter because $(\alpha^*, g^*) \in \mathcal{G}_{k+1}$, so that this pair contributes to the new models ϕ_{k+1}, ψ_{k+1}. Notice that ψ_k^* is a quadratic function. Expanding it at y^{k+1}, therefore gives

$$\psi_k^*(y,x) = \psi_k^*(y^{k+1},x) + \nabla\psi_k^*(y^{k+1},x)(y-y^{k+1}) + \tfrac{\delta}{2}(y-y^{k+1})^\top(y-y^{k+1}),$$

where $\nabla\psi_k^*(y,x) = g^* + \delta(y-x)$ and $\nabla^2\psi_k^*(y,x) = \delta I$. We now prove the formula

$$\psi_k^*(y,x) = \psi_k^*(y^{k+1},x) + \tfrac{\delta}{2}\|y-y^{k+1}\|^2. \tag{17}$$

Indeed, we have but to show that the first-order term in the above Taylor expansion vanishes at y^{k+1}. But this term is

$$\begin{aligned}
&\nabla\psi_k^*(y^{k+1},x)^\top(y-y^{k+1}) = \\
&\quad = \left[g^* + \delta(y^{k+1}-x)\right]^\top(y-y^{k+1}) \\
&\quad = g^{*\top}(y-y^{k+1}) + \delta(y^{k+1}-x)^\top(y-y^{k+1}) \\
&\quad = \delta(x-y^{k+1})^\top(y-y^{k+1}) + \delta(y^{k+1}-x)^\top(y-y^{k+1}) && \text{(using (9),(10))} \\
&\quad = 0,
\end{aligned}$$

and so formula (17) is established. Therefore

$$\begin{aligned}
\psi_k(y^{k+1},x) &\le \psi_k^*(y^{k+1},x) + \tfrac{\delta}{2}\|y^{k+2}-y^{k+1}\|^2 && \text{(using (16) left)} \\
&= \psi_k^*(y^{k+2},x) && \text{(using (17))} \\
&\le \psi_{k+1}(y^{k+2},x) && \text{(using (16) right)} \\
&\le \psi_{k+1}(x,x) && (y^{k+2} \text{ is minimizer of } \psi_{k+1}) \\
&= f(x).
\end{aligned}$$

This proves that the sequence $\psi_k(y^{k+1},x)$ is monotonically increasing and bounded above by $f(x)$, so it converges to some limit $\psi^* \le f(x)$. Since the term $\frac{\delta}{2}\|y^{k+2} - y^{k+1}\|^2$ is squeezed in between two terms with the same limit ψ^*, we deduce $\frac{\delta}{2}\|y^{k+2} - y^{k+1}\|^2 \to 0$. Since the sequence y^k is bounded, namely,

$$\|y^{k+1}\| \le \|x\| + \delta_1^{-1} \max_{\omega\in[0,\infty]} \|F'(x,\omega)^\star\|,$$

by formula (13), we deduce using a geometric argument that

$$\|y^{k+2}-x\|^2 - \|y^{k+1}-x\|^2 \to 0. \tag{18}$$

Recalling the relation $\phi_k(y,x) = \psi_k(y,x) - \frac{\delta}{2}\|y-x\|^2$, we finally obtain

$$\begin{aligned}
&\phi_{k+1}(y^{k+2},x) - \phi_k(y^{k+1},x) \\
&\qquad = \psi_{k+1}(y^{k+2},x) - \psi_k(y^{k+1},x) - \tfrac{\delta}{2}\|y^{k+2}-x\|^2 + \tfrac{\delta}{2}\|y^{k+1}-x\|^2 \to 0,
\end{aligned} \tag{19}$$

which converges to 0 due to convergence of $\psi_k(y^{k+1},x)$ proved above, and property (18).

3) Let e_{k+1} be the normalized eigenvector associated with the maximum eigenvalue of $F(x,\omega_{k+1}) + F'(x,\omega_{k+1})(y^{k+1}-x)$, which we pick in step 5 of the algorithm. Then $g_k = F'(x,\omega_{k+1})^* e_{k+1} e_{k+1}^\top$ is a subgradient of $\phi_{k+1}(\cdot,x)$ at y^{k+1}. That means

$$g_k^\top(y-y^{k+1}) \le \phi_{k+1}(y,x) - \phi_{k+1}(y^{k+1},x).$$

Using $\phi_{k+1}(y^{k+1},x) = \phi(y^{k+1},x)$ from Lemma 5 therefore implies

$$\phi(y^{k+1},x) + g_k^\top(y - y^{k+1}) \leq \phi_{k+1}(y,x). \tag{20}$$

Now observe that

$$\begin{aligned}
0 &\leq \phi(y^{k+1},x) - \phi_k(y^{k+1},x) \\
&= \phi(y^{k+1},x) + g_k^\top(y^{k+2} - y^{k+1}) - \phi_k(y^{k+1},x) - g_k^\top(y^{k+2} - y^{k+1}) \\
&\leq \phi_{k+1}(y^{k+2},x) - \phi_k(y^{k+1},x) + \|g_k\|\|y^{k+2} - y^{k+1}\| \qquad \text{(using (20))}
\end{aligned}$$

and this term tends to 0 because of (19), boundedness of g_k, and because $y^{k+1} - y^{k+2} \to 0$. We conclude that

$$\phi(y^{k+1},x) - \phi_k(y^{k+1},x) \to 0. \tag{21}$$

4) We now show that $\phi_k(y^{k+1},x) \to f(x)$, and therefore also $\phi(y^{k+1},x) \to f(x)$. Suppose on the contrary that $\eta := f(x) - \limsup \phi_k(y^{k+1},x) > 0$. Choose $0 < \theta < (1-\widetilde{\gamma})\eta$. It follows from (21) that there exists $k_1 \geq k_0$ such that

$$\phi(y^{k+1},x) - \theta \leq \phi_k(y^{k+1},x)$$

for all $k \geq k_1$. Using $\widetilde{\rho}_k < \widetilde{\gamma}$ for all $k \geq k_1$ gives

$$\begin{aligned}
\widetilde{\gamma}(\phi_k(y^{k+1},x) - f(x)) &\leq \phi(y^{k+1},x) - f(x) \\
&\leq \phi_k(y^{k+1},x) + \theta - f(x).
\end{aligned}$$

Passing to the limit implies $\widetilde{\gamma}\eta \geq \eta - \theta$, contradicting the choice of θ. This proves $\eta = 0$.

5) Having shown $\phi(y^{k+1},x) \to f(x)$, we now argue that we must have $y^{k+1} \to x$. This follows from the definition of y^{k+1}, because

$$\psi_k(y^{k+1},x) = \phi_k(y^{k+1},x) + \tfrac{\delta}{2}\|y^{k+1} - x\|^2 \leq \psi_k(x,x) = f(x).$$

Since $\phi_k(y^{k+1},x) \to f(x)$ by part 4), we have indeed $y^{k+1} \to x$. To finish the proof, observe that $0 \in \partial\psi_k(y^{k+1},x)$ implies

$$\begin{aligned}
\delta(x - y^{k+1})^\top(y - y^{k+1}) &\leq \phi_k(y,x) \ - \phi_k(y^{k+1},x) \\
&\leq \phi(y,x) - \phi_k(y^{k+1},x)
\end{aligned}$$

for every y. Passing to the limit implies

$$0 \leq \phi(y,x) - \phi(x,x),$$

because the left hand side converges to 0 in vue of $y^{k+1} \to x$. Since $\partial\phi(x,x) = \partial f(x)$, we are done. $\square$

8. Convergence of outer loop. All that remains to do now is piece things together and prove global convergence of our method. We have the following

THEOREM 8. *Suppose $x^1 \in \mathbb{R}^n$ is such that $\{x \in \mathbb{R}^n : f(x) \leq f(x^1)\}$ is compact. Then every accumulation point of the sequence x^j of serious iterates generated by our algorithm is a critical point of f.*

Proof. Let x^j be the sequence of serious steps. We have to show that $0 \in \partial f(\bar{x})$ for every accumulation point $\bar{x}$ of x^j. Suppose at the jth stage of the outer loop the inner loop accepts a serious step at $k = k_j$. Then $x^{j+1} = y^{k_j+1}$. By the definition of y^{k+1} as minimizer of the tangent program (8) this means

$$\delta_{k_j}\left(x^j - x^{j+1}\right) \in \partial\phi_{k_j}(x^{j+1}, x^j).$$

By convexity this can be re-written as

$$\delta_{k_j}\left(x^j - x^{j+1}\right)^\top \left(x^j - x^{j+1}\right) \leq \phi_{k_j}(x^j, x^j) - \phi(x^{j+1}, x^j) = f(x^j) - \phi_{k_j}(x^{j+1}, x^j),$$

the equality $\phi_{k_j}(x^j, x^j) = f(x^j)$ being true by Lemma 3. Since $x^{j+1} = y^{k_j+1}$ was accepted in step 4 of the algorithm, we have

$$f(x^j) - \phi_{k_j}(x^{j+1}, x^j) \leq \gamma^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Altogether

$$\delta_{k_j}\|x^j - x^{j+1}\|^2 \leq \gamma^{-1}\left(f(x^j) - f(x^{j+1})\right).$$

Summing over $j = 1, \dots, J-1$ gives

$$\sum_{j=1}^{J-1} \delta_{k_j}\|x^j - x^{j+1}\|^2 \leq \gamma^{-1}\sum_{j=1}^{J-1} f(x^j) - f(x^{j+1}) = \gamma^{-1}\left(f(x^1) - f(x^J)\right).$$

By hypothesis, f is bounded below on the set of iterates, because the algorithm is of descent type on the serious steps. This implies convergence of the series

$$\sum_{j=1}^{\infty} \delta_{k_j}\|x^j - x^{j+1}\|^2 < \infty.$$

In particular $\delta_{k_j}\|x^j - x^{j+1}\|^2 \to 0$. We claim now that $g_j = \delta_{k_j}\left(x^j - x^{j+1}\right) \to 0$.

Suppose on the contrary that there exists an infinite subsequence $j \in \mathcal{N}$ of $\mathbb{N}$ where $g_j = \delta_{k_j}\|x^j - x^{j+1}\| \geq \eta > 0$. Due to summability of $\delta_{k_j}\|x^j - x^{j+1}\|^2$ we must have $x^j - x^{j+1} \to 0$ in that case. That in turn is only possible when $\delta_{k_j} \to \infty$. We now construct another infinite subsequence $\mathcal{N}'$ of $\mathbb{N}$ such that $\delta_{k_j} \to \infty$, $j \in \mathcal{N}'$, and such that the doubling rule to increase δ_k in step 7 of the inner loop of the algorithm was applied at least once before $x^{j+1} = y^{k_j+1}$ was accepted. To construct $\mathcal{N}'$, we associate with every $j \in \mathcal{N}$ the last $j' \leq j$ where the δ-parameter was doubled while the inner loop was turning, and we let $\mathcal{N}'$ consists of all these j', $j \in \mathcal{N}$. It is possible that $j' = j$, but in general we can only assure that

$$2\delta_{k_{j'-1}} \leq \delta_{k_{j'}} \text{ and } \delta_{k_{j'}} \geq \delta_{k_{j'+1}} \geq \cdots \geq \delta_{k_j},$$

so that $\mathcal{N}'$ is not necessarily a subset of $\mathcal{N}$. What counts is that $\mathcal{N}'$ is infinite, that $\delta_{k_j} \to \infty$, $(j \in \mathcal{N}')$, and that the doubling rule was applied for each $j \in \mathcal{N}'$.

Let us say that at outer loop counter $j \in \mathcal{N}'$ it was applied for the last time in the inner loop at $\delta_{k_j-\nu_j}$ for some $\nu_j \geq 1$. That is, we have $\delta_{k_j-\nu_j+1} = 2\delta_{k_j-\nu_j}$, while the δ parameter was frozen during the remaining steps before acceptance in the inner loop, i.e.,

$$\delta_{k_j} = \delta_{k_j-1} = \cdots = \delta_{k_j-\nu_j+1} = 2\delta_{k_j-\nu_j}. \tag{22}$$

Recall from step 7 of the algorithm that we have $\rho_k < \gamma$ and $\tilde{\rho}_k \geq \tilde{\gamma}$ for those k where the step was not accepted and the doubling rule was applied. That is,

$$\rho_{k_j-\nu_j} = \frac{f(x^j) - f(y^{k_j-\nu_j+1})}{f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)} < \gamma$$

and

$$\tilde{\rho}_{k_j-\nu_j} = \frac{f(x^j) - \phi(y^{k_j-\nu_j+1}, x^j)}{f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)} \geq \tilde{\gamma}.$$

By (22) we now have

$$\tfrac{1}{2}\delta_{k_j}\left(x^j - y^{k_j-\nu_j+1}\right) \in \partial\phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j).$$

Using $\phi_{k_j-\nu_j}(x^j, x^j) = f(x^j)$ and the subgradient inequality for $\phi_{k_j-\nu_j}(\cdot, x^j)$ at $y^{k_j-\nu_j+1}$ gives

$$\begin{aligned}\tfrac{1}{2}\delta_{k_j}\left(x^j - y^{k_j-\nu_j+1}\right)^\top\left(x^j - y^{k_j-\nu_j+1}\right) &\leq \phi_{k_j-\nu_j}(x^j, x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)\\ &= f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j).\end{aligned}$$

This could also be written as

$$\frac{\delta_{k_j}\|x^j - y^{k_j-\nu_j+1}\|^2}{f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)} \leq 2. \tag{23}$$

Substituting (23) into the expression $\tilde{\rho}_{k_j-\nu_j}$ gives

$$\begin{aligned}\tilde{\rho}_{k_j-\nu_j} &= \rho_{k_j-\nu_j} + \frac{f(y^{k_j-\nu_j+1}) - \phi(y^{k_j-\nu_j+1}, x^j)}{f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)}\\ &\leq \rho_{k_j-\nu_j} + \frac{L\|x^j - y^{k_j-\nu_j+1}\|^2}{f(x^j) - \phi_{k_j-\nu_j}(y^{k_j-\nu_j+1}, x^j)}\\ &\leq \rho_{k_j-\nu_j} + \frac{2L}{\delta_{k_j}} \qquad \text{(using (23))}.\end{aligned}$$

Since $\rho_{k_j-\nu_j} < \gamma$ and $L/2\delta_{k_j} \to 0$, $(j \in \mathcal{N}')$, we have $\limsup_{j\in\mathcal{N}'} \tilde{\rho}_{k_j-\nu_j} \leq \limsup_{j\in\mathcal{N}'} \rho_{k_j-\nu_j} \leq \gamma$, contradicting $\tilde{\rho}_{k_j-\nu_j} \geq \tilde{\gamma} > \gamma$ for all $j \in \mathcal{N}'$.

Let us pick a convergent subsequence $x^j \to \bar{x}$, $j \in \mathcal{N}$. We wish to prove $0 \in \partial f(\bar{x})$. Observe that the sequence x^{j+1} is also bounded, so passing to a subsequence of $\mathcal{N}$ if necessary, we may assume $x^{j+1} \to \tilde{x}$, $j \in \mathcal{N}$. In general it could happen that $\tilde{x} \neq \bar{x}$. Only when δ_{k_j}, $j \in \mathcal{N}$, are bounded away from 0 can we conclude that $x^{j+1} - x^j \to 0$.

Now as $g_j = \delta_{k_j}(x^j - x^{j+1})$ is a subgradient of $\phi_{k_j}(\cdot, x^j)$ at $y^{k_j+1} = x^{j+1}$ we have

$$\begin{aligned}g_j^\top h &\leq \phi_{k_j}(x^{j+1} + h, x^j) - \phi_{k_j}(x^{j+1}, x^j)\\ &\leq \phi(x^{j+1} + h, x^j) - \phi_{k_j}(x^{j+1}, x^j) \qquad \text{(using } \phi_{k_j} \leq \phi\text{)}\end{aligned}$$

for every test vector h. Now we use the fact that $y^{k_j+1} = x^{j+1}$ was accepted in step 4 of the algorithm. That means

$$\gamma^{-1}\left(f(x^j) - f(x^{j+1})\right) \geq f(x^j) - \phi_{k_j}(x^{j+1}, x^j).$$

Combining these two estimates gives

$$\begin{aligned} g_j^\top h &\leq \phi(x^{j+1}+h, x^j) - f(x^j) + f(x^j) - \phi_{k_j}(x^{j+1}, x^j) \\ &\leq \phi(x^{j+1}+h, x^j) - f(x^j) + \gamma^{-1}\left(f(x^j) - f(x^{j+1})\right). \end{aligned}$$

Passing to the limit (using $g_j \to 0$, $x^{j+1} \to \tilde{x}$, $x^j \to \bar{x}$, $f(\bar{x}) = \phi(\bar{x};\bar{x})$, and $f(x^j) - f(x^{j+1}) \to 0$ in the order named) shows

$$0 \leq \phi(\tilde{x}+h;\bar{x}) - \phi(\bar{x};\bar{x})$$

for every h. This being true for every h, we can fix h' and choose $h = \bar{x} - \tilde{x} + h'$, which then gives

$$0 \leq \phi(\bar{x}+h';\bar{x}) - \phi(\bar{x};\bar{x}).$$

As this is true for every h', we have $0 \in \partial\phi(\cdot;\bar{x})(\bar{x})$, and hence also $0 \in \partial f(\bar{x})$. □

Acknowledgement. The authors acknowledge financial support from Agence Nationale de Recherche (ANR) under contract `SSIA_NV_6` *Controvert*, from Agence Nationale de Recherche (ANR) under contract `NT05 − 1_43040` *Guidage*, and from *Fondation d'Entreprise EADS* under contract *Solving challenging problems in feedback control*.

REFERENCES

[1] P. Apkarian, D. Noll, *Nonsmooth optimization for multidisk H_∞ synthesis.* European J. Control, to appear.
[2] P. Apkarian, D. Noll, *Controller design via nonsmooth multi-directional search.* SIAM J. Control and Optim., vol. 44, 2006, pp. 1923 - 1949.
[3] P. Apkarian, D. Noll, *IQC analysis and synthesis via nonsmooth optimization.* System and Control Letters, vol. 55, no. 12, 2006, pp. 971 – 981.
[4] P. Apkarian, D. Noll, *Nonsmooth H_∞ synthesis.* IEEE Trans. Autom. Control, vol. 51, 2006, pp. 71 - 86.
[5] P. Apkarian, D. Noll, *Nonsmooth optimization for multiband frequency domain control design.* Automatica, to appear.
[6] P. Apkarian, D. Noll, *Nonsmooth structured control design with applications to PID loop-shaping of a process.* International Journal of Robust and Nonlinear Control, to appear.
[7] P. Apkarian, D. Noll, O. Prot, *Nonsmooth methods for analysis and synthesis with integral quadratic constraints.* Submitted.
[8] P. Apkarian, D. Noll, O. Prot, *Trust region spectral bundle method for nonconvex maximum eigenvalue functions.* Submitted.
[9] P. Apkarian, D. Noll, J.-B. Thevenet and H.D. Tuan, *A spectral quadratic-SDP method with applications to fixed-order H_2 and H_∞ synthesis.* Asian Control Conference, Melbourne, AU, 2004.
[10] V. Bompart, P. Apkarian and D. Noll, *Nonsmooth techniques for stabilizing linear systems.* 2007 American Control Conference (2007 ACC), Times Square, New York, NY.
[11] V. Bompart, D. Noll and P. Apkarian, *Second-order nonsmooth optimization of the H_∞ norm.* Numerische Mathematik, to appear.
[12] J.F. Bonnas, J.-C. Gilbert, C. Lemaréchal and C. Sagastizábal, *A family of variable metric proximal methods.* Math. Programming, vol. 68, 1995, pp. 15 – 47.
[13] S. Boyd, V. Balakrishnan, *A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm.* System and Control Letters, vol. 15, 1990, pp. 1 - 7.
[14] S. Boyd, L. Elghaoui, E. Feron and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, vol. 15 of SIAM Studies in Applied Mathematics, SIAM, Philadelphia, 1994.
[15] B. M. Chen, *H_∞ control and its applications*, vol. 235 of Lecture Notes in Control and Information Sciences, Springer Verlag, New York, Heidelberg, Berlin, 1998.

[16] F. Clarke, *Optimization and nonsmooth analysis.* Canadian Math. Society Series, John Wiley & Sons, New York, 1983.

[17] J. Cullum, W. Donath and P. Wolfe, *The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices.* Math. Programming Studies, vol. 3, 1975, pp. 35 - 55.

[18] B. Fares, D. Noll and P. Apkarian, *Robust control via sequential semidefinite programming.* SIAM J. Control and Optim. vol. 40, 2002, pp. 1791 - 1820.

[19] R. Fletcher, *Semidefinite matrix constraints in optimization.* SIAM J. Control and Optim. vol. 23, 1985, pp. 493 - 513.

[20] A. Fuduli, M. Gaudioso and G. Giallombardo, *A DC piecewise affine model and a bundling technique in nonconvex nonsmooth optimization.* Optimization Method and Software, vol. 19, 2004, pp. 89 - 102.

[21] A. Fuduli, M. Gaudioso and G. Giallombardo, *Minimizing nonconvex nonsmooth functions via cutting planes and proximity control.* SIAM J. Optim. vol. 14, 2004, pp. 743 - 756.

[22] D. Gangsaas, K. Bruce, J. Blight and U.-L. Ly, *Applications of modern synthesis to aircraft control: Three case studies.* IEEE Trans. Autom. Control, AC-31, 1986, pp. 995 - 1014.

[23] C. Helmberg, K.C. Kiwiel, *A spectral bundle method with bounds.* Math. Programming, vol. 93, 2002, pp. 173 - 194.

[24] C. Helmberg, F. Oustry, *Bundle methods to minimize the maximum eigenvalue function.* Handbook of Semidefinite Programming. Theory, Algorithms and Applications. L. Vandenberghe, R. Saigal, H. Wolkowitz (eds.), vol. 27, 2000.

[25] C. Helmberg, F. Rendl, *Spectral bundle method for semidefinite programming.* SIAM J. Optimization, vol. 10, 2000, pp. 673 - 696.

[26] C. Helmberg, F. Rendl and R. Weismantel, *A semidefinite programming approach to the quadratic knapsack problem.* J. Combin. Optim. vol. 4, 2000, pp. 197 - 215.

[27] J.W. Helton, O. Merino, *Coordinate optimization for bi-convex matrix inequalities.* Proc. Conf. on Decis. Control, San Diego, CA, 1997, pp. 3609 - 3613.

[28] J.-B. Hiriart-Urruty, C. Lemaréchal , *Convex analysis and minimization algorithms, vol I: and vol II: Advanced theory and bundle methods*, vol. 306 of Grundlehren der mathematischen Wissenschaften, Springer Verlag, New York, Heidelberg, Berlin, 1993.

[29] Y.S. Hung, A.G.J. MacFarlane, *Multivariable feedback: A classical approach.* Lect. Notes in Control and Information Sciences, Springer Verlag, New York, Heidelberg, Berlin, 1982.

[30] L.H. Keel, S.P. Bhattacharyya and J.W. Howe, *Robust control with structured perturbations.* IEEE Trans Autom. Control vol. 36, 1988, pp. 68 - 77.

[31] K.C. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable optimization.* Math. Programming, vol. 46, 1990, pp. 105 - 122.

[32] K.C. Kiwiel, *Methods of descent for nondifferentiable optimization*, vol. 1133 of Springer Lect. Notes in Math., Springer Verlag, 1985.

[33] K.C. Kiwiel, *A linearization algorithm for computing control systems subject to singular value inequalities.* IEEE Trans. Autom. Control, AC-31, 1986, pp. 595 - 602.

[34] F. Leibfritz, *$COMPL_eIB$, COnstrained Matrix-optimization Problem LIbrary - a collection of test examples for nonlinear semidefinite programs, control system design and related problems.* Tech. Report, Universität Trier, 2003.

[35] C. Lemaréchal, *An extension of Davidson's method to nondifferentiable problems.* Math. Programming Studies, Nondifferentiable Optimization, M.L. Balinski and P. Wolfe (eds.), North Holland, 1975, pp. 95 - 109.

[36] C. Lemaréchal, *Bundle methods in nonsmooth optimization.* Nonsmooth Optimization, Proc. IIASA Workshop 1977, C. Lemaréchal, R. Mifflin (eds.), 1978.

[37] C. Lemaréchal, *Nondifferentiable optimization.* Chapter VII in: Handbooks in Operations Research and Management Sciences, vol. 1, 1989.

[38] C. Lemaréchal, *Lagrangian relaxation.* In: Computational and Combinatorial Optimization, M. Junger, D. Naddef (eds.), 2001, pp. 115 - 169.

[39] C. Lemaréchal, A. Nemirovskii and Y. Nesterov, *New variants of bundle methods.* Math. Programming, vol. 69, 1995, pp. 111 - 147.

[40] C. Lemaréchal, F. Oustry, *Nonsmooth algorithms to solve semidefinite programs.* SIAM series; Advances in Linear Matrix Inequalitiy Methods in Control, L. Elghaoui, S.-N. Niculescu (eds.), 2000.

[41] C. Lemaréchal, C. Sagastizábal, *Variable metric bundle methods: from conceptual to implementable forms.* Math. Programming, vol. 76, 1997, pp. 393 – 410.

[42] D. Mayne, E. Polak, *Algorithms for the design of control systems subject to singular value inequalities.* Math. Programming Studies, vol. 18, 1982, pp. 112 - 134.

[43] D. Mayne, E. Polak and A. Sangiovanni, *Computer aided design via optimization.* Auto-

matica, vol. 18, no. 2, 1982, pp. 147 - 154

[44] K. Mombaur, *Stability optimization of open-loop controlled walking robots*. PhD Thesis, Universität Heidelberg, 2001.

[45] D. Noll, P. Apkarian, *Spectral bundle method for nonconvex maximum eigenvalue functions: first-order methods*. Math. Programming, Series B, vol. 104, 2005, pp. 701 - 727.

[46] D. Noll, P. Apkarian, *Spectral bundle method for nonconvex maximum eigenvalue functions: second-order methods*. Math. Programming, Series B, vol. 104, 2005, pp. 729 - 747.

[47] D. Noll, M. Torki and P. Apkarian, *Partially augmented Lagrangian method for matrix inequality constraints*. SIAM J. on Optimization, vol. 15, 2004, pp. 161 - 184.

[48] F. Oustry, *A second-order bundle method to minimize the maximum eigenvalue function*. Math. Programming, Series A, vol. 89, 2000, pp. 1 - 33.

[49] M. Overton, *On minimizing the maximum eigenvalue of a symmetric matrix*. SIAM J. on Matrix Analysis and Applications, vol. 9, 1988, pp. 256 - 268.

[50] M. Overton, *Large-scale optimization of eigenvalues*. SIAM J. on Optimization, vol. 2, 1992, pp. 88 - 120.

[51] M. Overton, R. Womersley, *On the sum of the largest eigenvalues of a symmetric matrix*. SIAM J. on Matrix Analysis and Applications, vol. 13, 1992, pp. 41 - 45.

[52] E. Polak, *On the mathematical foundations of nondifferentiable optimization in engineering design*. SIAM Review, vol. 29, 1987, pp. 21 - 89.

[53] E. Polak, *Optimization: Algorithms and Consistent Approximations*. Springer Series in Applied Mathematical Sciences, vol. 124, 1997.

[54] E. Polak, Y. Wardi, *A nondifferential optimization algorithm for the design of control systems subject to singular value inequalities over the frequency range*. Automatica, vol. 18, 1982, pp. 267 - 283.

[55] F. Pukelsheim, *Optimal design of experiments*. John Wiley & Sons, New York, 1993.

[56] J.-B. Thevenet, D. Noll and P. Apkarian, *Nonlinear spectral SDP method for BMI-constrained problems: applications to control design*. Informatics in Control, Automation and Robotics I, J. Braz, H. Arajo A. Viera and B. Encarnaco (eds.), Springer Verlag, 2006, pp. 61 - 72.

[57] J.-B. Thevenet, D. Noll and P. Apkarian, *Nonsmooth methods for large bilinear matrix inequalities: applications to feedback control*. Submitted.

[58] L. Vandenberghe, S. Boyd, *Semidefinite programming*. SIAM Review, vol. 38, 1996, pp. 49 - 95.

[59] P. Wolfe, *A method of conjugate subgradients for minimizing nondifferentiable functions*. Math. Programming Studies, vol. 3. Nondifferentiable Optimization, M.L. Balinski, P. Wolfe (eds.), North-Holland, 1975, pp. 145 - 173.

[60] Z. Zhao, B. Braams, M. Fukuda, M. Overton and J. Percus, *The reduced density matrix for electronic structure calculations and the role of three-index representability*. J. Chem. Physics, vol. 11, 2003.