Computer Sciences Department
University of Wisconsin
1210 West Dayton Street
Madison, Wisconsin 53706

A SIMPLE, GLOBAL, COMPLEMENTARY VARIATIONAL PRINCIPLE*

by

James W. Daniel

Technical Report #63

June 1969

*Prepared under Contract Number N00014-67-A-0128-0004 at the University of Wisconsin. Reporduction in whole or in part is permitted for any purpose of the United States Government.

1. INTRODUCTION

This brief note presents a simple, global, complementary variational principle for a broad class of general functions. That is, we show that a point $\hat{u}$ which minimizes a certain type of functional f over a set S also maximizes a computable functional f^* over a set S^* and satisfies $f(\hat{u}) = f^*(\hat{u})$; this fact can be used to provide error bounds for an approximate minimizer u of f.

Complementary principles have been studied rather thoroughly for differential equations [5, 6, references therein] and general operator extensions of the differential equation problem [7], although primarily local principles from a somewhat different viewpoint are obtained. A global principle has been presented in [2, 9] for nonlinear boundary value problems of variational type in order to obtain variational error bounds; this present note extends those results to more general functionals.

2. THE MAIN THEOREM

Theorem 1. Let f be a real valued nonlinear functional on a real Hilbert space H with inner product $\langle \cdot , \cdot \rangle$, and let $\hat{u}$ minimize f on the norm closed convex set S. For each w in some given set $S^* \subset H$ let a self-adjoint linear operator P_w be defined, satisfying $\langle P_w h, h \rangle \geq a_w \langle h, h \rangle$ for all h in H, with $a_w > 0$. Suppose that f is twice continuously Frechet differentiable on the set of points of the form $\lambda v + (1-\lambda)w$ for v in S,

w in S^*, and λ in $[0,1]$, and that $\langle [f''_{\lambda v+(1-\lambda)w} - P_w](v-w), v-w\rangle \geq 0$ for all v in S, w in S^*, and λ in $[0,1]$. For each w in S^*, let v_w in S be defined as the (unique) point in S minimizing $\frac{1}{2}\langle P_w v, v\rangle + \langle \nabla f(w) - P_w w, v\rangle$, and let f^* be defined on S^* as

$$f^*(w) = f(w) + \tfrac{1}{2}\langle v_w - w, P_w(v_w - w)\rangle + \langle v_w - w, \nabla f(w)\rangle .$$

If $\hat{u}$ is in S^*, then $f^*(w)$ is maximized by $w = \hat{u}$, and $f^*(\hat{u}) = f(\hat{u})$. For all u in S and w in S^* we have the estimates

$$f(u) - f^*(w) \geq \tfrac{1}{2}\langle P_w(\hat{u} - v_w), \hat{u} - v_w\rangle$$

$$f(u) - f^*(w) \geq \tfrac{1}{2}\langle P_{\hat{u}}(\hat{u} - u), \hat{u} - u\rangle .$$

<u>Proof</u>: Since $\hat{u}$ minimizes f over the convex set S, we have $\langle \nabla f(\hat{u}), s - \hat{u}\rangle \geq 0$ for all s in S. Let $g(v) = \frac{1}{2}\langle P_{\hat{u}} v, v\rangle + \langle \nabla f(\hat{u}) - P_{\hat{u}}\hat{u}, v\rangle$. Then $\langle s-\hat{u}, \nabla g(\hat{u})\rangle = \langle s-\hat{u}, P_{\hat{u}}\hat{u} + \nabla f(\hat{u}) - P_{\hat{u}}\hat{u}\rangle = \langle s-\hat{u}, \nabla f(\hat{u})\rangle \geq 0$ for all s in $\hat{u}$. Since the quadratic functional $g(v)$ has $g''_v \equiv P_{\hat{u}}$, a positive definite operator, a unique point $v_{\hat{u}}$ exists (similarly for v_w for all w in S^*) and is characterized by the condition $\langle s - v_{\hat{u}}, \nabla g(v_{\hat{u}})\rangle \geq 0$; therefore $v_{\hat{u}} = \hat{u}$ and hence $f^*(\hat{u}) = f(\hat{u})$. To prove that $\hat{u}$ maximizes f^*, we write, for w in S^*, letting $v_w \equiv v$ and $P_w \equiv P$ for notational ease,

$$f^*(\hat{u}) - f^*(w) = f(\hat{u}) - f(w) - \tfrac{1}{2}\langle v-w, P(v-w)\rangle - \langle v-w, \nabla f(w)\rangle$$

$$= \tfrac{1}{2}\langle P(\hat{u}-v), \hat{u}-v\rangle + [f(\hat{u}) - \tfrac{1}{2}\langle P\hat{u}, \hat{u}\rangle]$$

$$- [f(w) - \tfrac{1}{2}\langle Pw, w\rangle] + \langle \hat{u}-v, Pv-Pw+\nabla f(w)\rangle$$

$$+ \langle \hat{u}-w, Pw - \nabla f(w)\rangle$$

by adding and subtracting $\frac{1}{2}\langle P(\hat{u}-v), \hat{u}-v\rangle$. Since v minimizes $\frac{1}{2}\langle Pv, v\rangle + \langle \nabla f(w) - Pw, v\rangle$ over S and $\hat{u}$ is in S, the inner product inequality characterizing v gives $\langle \hat{u}-v, Pv-Pw+\nabla f(w)\rangle \geq 0$. Hence

$$f^*(\hat{u}) - f^*(w) \geq \tfrac{1}{2}\langle P(\hat{u}-v), \hat{u}-v\rangle + d(\hat{u}) - d(w) + \langle \hat{u}-w, Pw - \nabla f(w)\rangle$$

where $d(u) \equiv f(u) - \frac{1}{2}\langle Pu, u\rangle$. We have

$$d(\hat{u}) - d(w) = \langle \nabla d(w), \hat{u}-w\rangle + \int_0^1 t\,\langle d''_{t\hat{u}+(1-t)w}(\hat{u}-w), \hat{u}-w\rangle\,dt$$

$$= \langle \nabla f(w) - Pw, \hat{u}-w\rangle + \int_0^1 t\,\langle [f''_{t\hat{u}+(1-t)w} - P_w](\hat{u}-w), \hat{u}-w\rangle\,dt$$

$$\geq \langle \nabla f(w) - Pw, \hat{u}-w\rangle .$$

Inserting this we find

$$f^*(\hat{u}) - f^*(w) \geq \tfrac{1}{2}\langle P(\hat{u}-v), \hat{u}-v\rangle + \langle \nabla f(w) - Pw, \hat{u}-w\rangle + \langle \hat{u}-w, Pw - \nabla f(w)\rangle$$

$$\geq \tfrac{1}{2}\langle P_w(\hat{u}-v), \hat{u}-v\rangle \geq \frac{a_w}{2}\,\|\hat{u}-v\|^2 .$$

Thus $\hat{u}$ maximizes f^* over S^*. To obtain the first error estimate, we merely write $f(u) - f^*(w) \geq f(\hat{u}) - f^*(w) = f^*(\hat{u}) - f^*(w) \geq \frac{1}{2}\langle P_w(\hat{u}-v_w), \hat{u}-v_w\rangle$.

For the second, we write

$$f(u) - f^*(w) \geq f(u) - f^*(\hat{u}) = f(u) - f(\hat{u}) = \langle u-\hat{u}, \nabla f(\hat{u})\rangle$$

$$+ \int_0^1 t \langle [f''_{tu+(1-t)\hat{u}} - P_{\hat{u}}](u-\hat{u}), u-\hat{u}\rangle dt + \tfrac{1}{2}\langle P_{\hat{u}}(u-\hat{u}), u-\hat{u}\rangle$$

$$\geq \tfrac{1}{2}\langle P_{\hat{u}}(u-\hat{u}), u-\hat{u}\rangle$$

by the assumption on $f'' - P$ and the necessary condition for $\hat{u}$ to minimize f over S, since u is in S.

Q.E.D.

Remarks:

i) The closedness and convexity of S are used only to guarantee the existence and uniqueness of v_w for w in S and to deduce that $v_{\hat{u}} = \hat{u}$; these properties can be guaranteed in other ways as well.

ii) The differentiability hypotheses can be weakened easily; in particular, $f(v) - \frac{1}{2}\langle P_w v, v\rangle$ need only be differentiably convex on S for each w in S^*.

iii) Independent of any convexity hypotheses on f, points other that $\hat{u}$ can maximize f^* without added restrictions on $f'' - P$; this is of no consequence for our pruposes of error bounding, however.

iv) In order for the error bounds to be effective, one would require that f^* is continuous at $\hat{u}$; this requires further study of P_w. An examination of the expression for $f^*(\hat{u}) - f^*(w)$ shows that if f''_u and P_w are uniformly bounded for u and w near $\hat{u}$ then for some constants a, b, c we have $0 \leq f^*(\hat{u}) - f^*(w) \leq a\|\hat{u} - v_w\|^2 + b\|\hat{u} - w\|^2 + c\|\hat{u} - v_w\| \; \|\hat{u} - w\|$;

thus we need only study $\hat{u} - v_w = v_{\hat{u}} - v_w$ for w near $\hat{u}$. If for example $a_w \geq \varepsilon > 0$, $\| \nabla f(w) - \nabla f(\hat{u}) \| \leq K \| w - \hat{u} \|$ near $\hat{u}$, and $S = H$, then $\| \hat{u} - v_w \| \leq (1 + \frac{K}{\varepsilon}) \| \hat{u} - w \|$. For the general situation we have the following more restrictive situation.

Theorem 2. Let the assumptions of Theorem 1 hold. Moreover suppose that $\| f''_w \| \leq A$, $\| P_w - P_{\hat{u}} \| \leq A \| w - \hat{u} \|$, and $a_w \geq \varepsilon > 0$, all for w near $\hat{u}$. Then $\| \hat{u} - v_w \| = O(\| \hat{u} - w \|^{\frac{1}{2}})$ and hence $f^*(\hat{u}) - f^*(w) = O(\| \hat{u} - w \|)$.

Proof: Let $g_w(v) \equiv \frac{1}{2} \langle P_w v, v \rangle + \langle \nabla f(w) - P_w w, v \rangle$. Then $| g_{\hat{u}}(v) - g_w(v) | = O(\| v \| \; \| \hat{u} - w \|)$. Then we have $g_w(v_w) \leq g_w(\hat{u}) = g_{\hat{u}}(\hat{u}) + [g_w(\hat{u}) - g_{\hat{u}}(\hat{u})] \leq g_{\hat{u}}(\hat{u}) + O(\| \hat{u} \| \; \| \hat{u} - w \|)$ and similarly $g_{\hat{u}}(\hat{u}) \leq g_{\hat{u}}(v_w) \leq g_w(v_w) + O(\| v_w \| \; \| \hat{u} - w \|)$. Hence $| g_{\hat{u}}(\hat{u}) - g_w(v_w) | \leq \| v_w \| O(\| \hat{u} - w \|)$. Now also for w near $\hat{u}$ $g_w(v) \geq \frac{1}{2} a_w \| v \|^2 - \| v \| \cdot M$ for a fixed constant M, and since $g_w(0) = 0$ we have $\| v_w \| \leq \frac{2M}{\varepsilon}$. In addition $g_{\hat{u}}(v) = g_{\hat{u}}(\hat{u}) + \langle \nabla g_{\hat{u}}(\hat{u}), v - \hat{u} \rangle + \frac{1}{2} \langle P_{\hat{u}}(v - \hat{u}), v - \hat{u} \rangle \geq g_{\hat{u}}(\hat{u}) + \frac{\varepsilon}{2} \| v - \hat{u} \|^2$. Therefore $\frac{\varepsilon}{2} \| v_w - \hat{u} \|^2 \leq g_{\hat{u}}(v_w) - g_{\hat{u}}(\hat{u}) \leq g_w(v_w) + O(\| v_w \| \; \| \hat{u} - w \|) - g_{\hat{u}}(\hat{u}) \leq O(\| \hat{u} - w \|) + g_w(v_w) - g_{\hat{u}}(\hat{u}) \leq O(\| \hat{u} - w \|)$.

Q.E.D.

3. EXAMPLES

We wish to give two concrete examples to the meaning of the general theorem; it is simplest to consider differential equations, and in order to minimize technical complexities we consider the equation

$$u''(t) = c(t, u(t)), \quad t \text{ in } (0, 1), \quad u(0) = u(1) = 0 .$$

More precisely we consider minimizing the functional

$$f(u) \equiv \tfrac{1}{2} \int_0^1 [u'(t)]^2 \, dt + \int_0^1 \int_0^{u(t)} c(t, x) dx \, dt$$

over the set H of absolutely continuous functions u having u' in $L_2(0, 1)$ and $u(0) = u(1) = 0$. For u, v in H, we take $\langle u, v\rangle = \int_0^1 [u'(t)v'(t) + u(t)v(t)]dt$. Since we wish to illustrate ideas rather that technicalities in this section, we shall be rather sloppy and speak blithely of $D^2 u \equiv u''$ for u in H; the precise formulation is easily filled in.

A) For the first example, let us suppose that $c_u(t, u) \geq \gamma > -\pi^2$ for all u in $(-\infty, \infty)$, t in $[0, 1]$. Let $(u, v) = \int_0^1 u(t)\, v(t)\, dt$. Then $f(u+h) = f(u) + (-D^2 u + c(t, u), h) + \frac{1}{2} ([-D^2 + c_u(t, u)]\, h, h) +$ small terms. Thus $\langle \nabla f(u), h\rangle = (-D^2 u + c(t, u), h)$ and $\langle f''_u h, h\rangle = ([-D^2 + c_u(t, u)]h, h)$. We let $S = S^* = H$ and for all w define P_w by $-D^2 + \gamma$ which is positive definite since $\gamma > -\pi^2$. Since $c_u \geq \gamma$, we have $\langle [f''_u - P_w]h, h\rangle = ([c_u - \gamma]h, h) \geq 0$ and the hypotheses are fulfilled. Here $f^*(w) = \frac{1}{2}\int_0^1 [w'(t)]^2 dt + \int_0^1 \{\frac{\gamma}{2} v^2(t) + (v(t) - w(t))[c(t, w(t)) - \gamma\, w(t)] + \int_0^{w(t)} (c(t, x) - \gamma x)\, dx\} dt$ where $v(t)$ solves

$$v''(t) - \gamma v(t) = c(t, w(t)) - \gamma w(t) \text{ for } t \text{ in } (0, 1), v(0)=v(1) = 0 .$$

In this case, discussed in [9], the error bounds are in the norm $\langle Pe, e\rangle = \int_0^1 \{[e'(t)]^2 + \gamma\, e^2(t)\}dt$. This yields useful bounds for any approximate solution w and for the corresponding v_w . Such a w might for example be obtained by the Ritz procedure or by an iterative process. For some problems the Newton iterative process yields a sequence u_n decreasing to the desired solution [1, 2, 3, 8]; often then v_{u_n} turns out to lie below the solution [8] yielding error bounds for $\hat{u}$. The variational procedure above in addition furnishes bounds involving the derivatives.

B) In some cases, the Newton iteration mentioned above may be costly to carry out. Certain Picard type iterations, though more slowly convergent, are sometimes used at least until one is near the solution where Newton's method might be worth the cost. The process above in A) will often yield two-sided bounds and the $\langle P_e e, e\rangle$ bounds as well in this case too. We wish to observe than one Newton step also provides such bounds in some cases. Suppose now that $|c(t, u)| \le N$ for all t, u, that u_0 solves $u_0'' = -N$, $u_0(0) = u_0(1) = 0$, that $k \ge c_u(t, u) \ge \gamma > -\pi^2$. Then it is known [4] that the sequence

$$u''_{n+1} - k\, u_{n+1} = c(t, u_n) - k\, u_n, \quad u_n(0) = u_n(1) = 0, \quad n = 0, 1, \ldots$$

is a monotone decreasing sequence converging to the solution $\hat{u}$. Now let $S = \{u; u \le \hat{u} \text{ in } [0, 1]\}$, $S^* = \{u; u \ge \hat{u} \text{ in } [0, 1]\}$. For w in S^*, define P_w by $-D^2 + c_w(t, w)$. As we saw before this is positive definite. Let us also suppose that $\gamma \ge 0$, that is $c_u(t, u) \ge 0$, and $c_{uu}(t, u) \le 0$. Then for w in S^* and u in S we have $\langle [f''_{w+\lambda(u-w)} - P_w](u-w), u-w\rangle =$ $([c(t, w+\lambda(u-w)) - c(t, w)](u-w), u-w) \ge 0$ since $u \le w$ and $c_{zz} \le 0$ implies $c(t, w+\lambda(u-w)) \ge c(t, w)$. Thus the hypotheses are satisfied. We now claim that, for w in S^*, the v_w that minimizes $\frac{1}{2}\langle v, P_w v\rangle + \langle \nabla f(w) - P_w w, v\rangle$ over S in fact minimizes it over all H, that is, that the gradient

$$P_w v + \nabla f(w) - P_w w = 0 \text{ at } v = v_w ;$$

this is well known. To do this, we show that if $P_w v + \nabla f(w) - P_w v = 0$ then v is in S and hence $v = v_w$. This equation for v yields

$$v'' - c_w(t,w)v = c(t,w) - c_w(t,w)w, \quad v(0) = v(1) = 0$$

which is just the Newton iteration from w to v. Since also

$$\hat{u}'' - c_w(t,w)\,\hat{u} = c(t,\hat{u}) - c_w(t,w)\,\hat{u}, \quad \hat{u}(0) = \hat{u}(1) = 0 ,$$

subtracting we have

$$(v-\hat{u})'' - c_w(t,w)(v-\hat{u}) = c(t,w) - c(t,\hat{u}) - c_w(t,w)(w-\hat{u}) \geq 0$$

since $w \geq \hat{u}$ and $c_{uu} \leq 0$. But then the maximum principle implies that $v-\hat{u} \leq 0$, that is, that v is in S and hence $v = v_w$. Thus we find

$$f^*(w) = \frac{1}{2}\int_0^1 [v'(t)]^2 dt + \int_0^1\int_0^{w(t)} c(t,x)dx\,dt + \frac{1}{2}\int_0^1 c_w(t,w(t))(v(t)-w(t))^2 dt$$

$$+ \int_0^1 c(t,w(t))\,(v(t) - w(t))\,dt$$

where v is the Newton iterate of w solving

$$v'' - c_w(t,w)\,v = c(t,w) - c_w(t,w)w, \quad v(0) = v(1) = 0 .$$

These error bounds are in the norms $\int_0^1 \{e'(t)]^2 + c_z(t,z)\,e^2(t)\}\,dt$ for $z = w$ and $z = \hat{u}$. Since $c_z(t,z) \geq 0$, we have bounds for $\int_0^1 [e'(t)]^2\,dt$ as well as the fact that $v_w \leq \hat{u} \leq w$. The computable bound for derivatives would thus use the variational results, $\frac{1}{2}\int_0^1 [\hat{u}'(t) - v_w'(t)]^2\,dt \leq f(v_w) - f(w)$.

4. REFERENCES

1. Bellman, R., "On monotone convergence to solutions of $u' = g(u, t)$", Proc. Amer. Math. Soc., vol. 8 (1957), 1007-1009.

2. Bellman, R., "Quasi-linearization and upper and lower bounds for variational problems," Quart. Appl. Math., vol. 19, (1962), 349-350

3. Collatz, L., Functional analysis and applied mathematics, Academic Press, New York, 1966.

4. Courant, R., Hilbert, D., Methods of mathematical physics, Vol. II, Interscience, New York, 1962.

5. Noble, B., "Complementary variational principles for boundary value problems I: Basic principles," Math. Res. Cntr. Rept. # 473, (1964).

6. Noble, B., "Complementary variational principles II: nonlinear networks," Math. Res. Cntr. Rept. # 643 (1966).

7. Rall, L., "On complementary variational principles," J. Math. Anal. Appl., Vol. 14 (1966), 174-184.

8. Shampine, L., "Monotone iterations and two-sided convergence," SIAM J. Num. Anal., vol. 3 (1966), 607-615.

9. Shampine, L., "Error bounds and variational methods for nonlinear boundary value problems," Numer. Math., Vol. 12 (1968), 410-415.