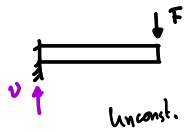
⑤ Constrained optimization in Computational engeenering

* Interpretation of Lagrange multipliers in computational engeenering

a) Dirichlet boundary conditions



$$\min_{u} \ \tfrac{1}{2} u^T K u - f^T u$$
$$\text{s.t.} \quad u = 0 \ \text{ in } \Gamma_D \ (\nu)$$

Unconst.

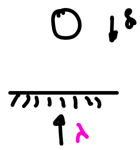
if no B.C. $\min \tfrac{1}{2} u^T K u - f^T u$

B.C.: $Ku = f$

Constr. $\max_{\nu} \min_{u} \ \overbrace{\tfrac{1}{2} u^T K u - f^T u + \nu^T u}$

KKT: $\begin{bmatrix} K & I \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ \nu \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}$

b) Contact problem



$$\min_{u} \ \tfrac{1}{2} \dot{u}^T M \dot{u} + mg u$$
$$u \geq 0 \ (\lambda)$$

c) Stokes problem

$$\min \int \nu |\nabla u|^2$$
$$\text{s.t.} \quad \nabla \cdot u = 0 \ (p)$$

$$\min \tfrac{1}{2} \nu \int |\nabla u|^2$$
$$\text{s.t.} \quad \int q (\nabla \cdot u) = 0 \quad \forall q$$

$$L(u, p) = \tfrac{1}{2} \nu \int |\nabla u|^2 - \int p \nabla \cdot u$$

$$\max_{p} \min_{u} \ \underbrace{\nu \int |\nabla u|^2 - \int p \nabla \cdot u}$$

KKT
$$\begin{cases} \dfrac{\partial L}{\partial u}(v) = \nu \int \nabla u \nabla v - \int p \nabla \cdot v = 0 \quad \forall v \\ \dfrac{\partial L}{\partial p}(q) = \int q \nabla \cdot u = 0 \quad \forall q \end{cases}$$

KKT (Euler-Lagrange)
$$\begin{cases} \nu \int \nabla u \nabla v - \int p \nabla u = 0 \\ \int q \nabla \cdot u = 0 \end{cases}$$

$$\begin{cases} -\Delta u + p = 0 \\ \nabla \cdot u = 0 \end{cases}$$

$$\begin{bmatrix} K & D^T \\ D & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}$$

d) Eigenvalue problems

$$\|A\|_2 := \sup_{u} \frac{|Au|_2}{\|u\|_2} = \sup_{u} \frac{u^T A^T A u}{u^T u}$$

$$u^* = \arg\sup_{u} \frac{u^T A^T A u}{u^T u} = \left\{ \begin{array}{l} \arg\sup_{u} \ \tfrac{1}{2} u^T A^T A u \\ \text{s.t.} \quad u^T u = 1 \end{array} \right\} \underset{\bar{A} = A^T A}{=} \left\{ \begin{array}{l} \arg\min_{u} -\tfrac{1}{2} u^T \bar{A} u \\ \text{s.t.} \quad u^T u = 1 \ (\nu) \end{array} \right\} \overset{\text{KKT}}{\Rightarrow} \left\{ \begin{array}{l} \bar{A} u - \nu u = 0 \\ u^T u = 1 \end{array} \right\}$$

$$\sup_{u} \frac{u^T A^T A u}{u^T u} = \left\{ \begin{array}{l} \sup_{u,\, c>0} \ \dfrac{u^T A^T A u}{c^2} \\ \text{s.t.} \quad u^T u = c^2 \end{array} \right\} = \left\{ \begin{array}{l} \sup_{u,\, c>0} \ \frac{u^T}{c} A^T A \frac{u}{c} \\ \text{s.t.} \quad \frac{u^T}{c} \frac{u}{c} = 1 \end{array} \right\} = \left\{ \begin{array}{l} \sup_{\omega} \ \omega^T A^T A \omega \\ \text{s.t.} \quad \omega^T \omega = 1 \end{array} \right\}$$

e) Plasticity : { Maximum dissipation
Yield surface

f) Schur complement: { Solving the dual in quadratic problems

Interior point method

$$\begin{cases} \min & f_0(\theta) \\ \text{s.t} & f_i(\theta) \leq 0 \quad (\lambda) \\ & A\theta = b \quad (\nu) \end{cases}$$

<u>KKT</u>:
$$\left\{ \begin{array}{l} A\theta^* = b \\ f_i(\theta^*) \leq 0 \\ \lambda^* \geq 0 \\ \nabla f_0(\theta^*) + \lambda_i^* \nabla f_i(\theta^*) + \underbrace{A^T}_{\nabla h} \nu = 0 \\ \lambda_i^* f_i(\theta^*) = 0 \end{array} \right\}$$

	Optim cond	Algorithms
Uncont. opt:	$\nabla f_0(\theta) = 0$	* Gradient method * Newton * Quasi-newt.
Constr opt.	KKT	* Augm. Lagr. * Interior point

ADMM
Nesterov accel.
Proximal method

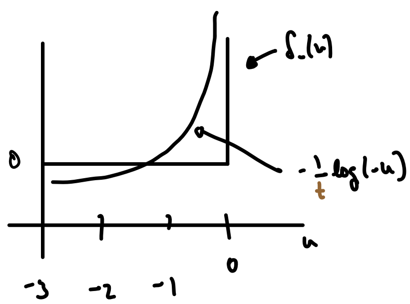
* Logarithmic barrier

Recall:

$$\begin{cases} \min & f_0(\theta) \\ \text{s.t} & f_i(\theta) \leq 0 \quad (\lambda) \\ & A\theta = b \quad (\nu) \end{cases} \Rightarrow \begin{cases} \min & f_0(\theta) + \delta_-(f_i(\theta)) \\ \text{s.t} & A\theta = b \quad (\nu) \end{cases}$$

smooth approximation

$$\begin{cases} \min & f_0(\theta) - \frac{1}{t} \log(-f_i(\theta)) \\ \text{s.t} & A\theta = b \quad (\nu) \end{cases}$$



<u>New KKT</u>:
$$\left\{ \begin{array}{l} A\theta^* = b \\ f_i(\theta^*) \leq 0 \\ \lambda^* \geq 0 \\ \nabla f_0(\theta^*) + \lambda_i^* \nabla f_i(\theta^*) + A^T \nu = 0 \\ -\lambda_i^* f_i(\theta^*) = 1/t \end{array} \right\}$$

* Remark:

- Use Newton iterations for each t
- Make t increase during iterations
- In convex problem, it can be shown that iterations travel in the central path
- IPOPT is an open-source library for large constr. optim.

Augmented Lagrangian

* Introduction. Let's focus on linear constraints.

Base problem

$$(B)\begin{cases} \min\limits_{\theta} \; f_0(\theta) \\ \text{s.t.} \;\; h(\theta)=0 \;\; (\nu) \end{cases}$$

$$\text{KKT}: \left\{ \begin{array}{l} \nabla f_0(\theta)+\nu\nabla h(\theta)=0 \\ h(\theta)=0 \end{array} \right\} \Rightarrow F(\theta,\nu)=0$$

Quadratic penalization problem

$$(QP)\left\{ \min\limits_{\theta} \; f(\theta)+\tfrac{1}{2}\rho h(\theta)^2 \right.$$

optimality cond:

$$\nabla f(\theta)+\rho h(\theta)\nabla h(\theta)=0$$

Linear penalization problem

$$(LP)\begin{cases} \min\limits_{\theta} \; f(\theta)+\nu_k h(\theta) \\ \text{s.t.} \;\; h(\theta)=0 \;\; (\nu) \end{cases}$$

$$\text{KKT}: \left\{ \begin{array}{l} \nabla f_0(\theta)+(\nu+\nu_k)\nabla h(\theta)=0 \\ h(\theta)=0 \end{array} \right\} \Rightarrow F(\theta,\nu)=0$$

Quadratic and linear penalization problem

$$(QLP)\left\{ \min\limits_{\theta} \; f(\theta)+\nu_k h(\theta)+\tfrac{1}{2}\rho h(\theta)^2 \right.$$

optimality cond:

$$\nabla f(\theta)+[\nu_k+\rho h(\theta)]\nabla h(\theta)=0$$

* **Regularization of the base problem**

$$(RP)\begin{cases} \min \; f(\theta)+\tfrac{1}{2}\rho h(\theta)^2 \\ \text{s.t.} \;\; h(\theta)=0 \;\; (\nu) \end{cases}$$

optimality cond:

$$\left. \begin{array}{l} \nabla f(\theta)+[\nu+\rho h(\theta)]\nabla h(\theta)=0 \\ h(\theta)=0 \end{array} \right\} \Rightarrow F(\theta,\nu)=0$$

Constrained | Unconstrained

$$\begin{tikzcd} (B) \arrow[r, "\text{Eliminating dof } \nu"] \arrow[d, "\text{Adding dofs } \nu_k"'] & (QP) \arrow[d] \\ (LP) \arrow[r] & (QLP) \end{tikzcd}$$

ρ = penalty parameter

* Remark ①: If $\nu_k=\nu^*$, then (QLP) as an exact quadratic penalization of (B) problem

* If $\nu_k+\rho h(\theta^*)=\nu^*$, then the optimality conditions of (QLP) / (RP) coincides with the ones of (B)

* The algorithm:

a) $\theta_{k+1} = \arg\min_{\theta} \{ L(\theta) + \nu_k h(\theta) + \frac{1}{2}\rho\, h(\theta)^2$

b) $\nu_{k+1} = \nu_k + \rho\, h(\theta_{k+1})$

* Dual view point

(B) $\begin{cases} \min_{\theta} & f_0(\theta) \\ \text{s.t.} & h(\theta) = 0 \quad (\nu) \end{cases}$

Lagrangian: $L(\theta, \nu) = f_0(\theta) + \nu^T h(\theta)$

Dual function: $g(\nu) = \min_{\theta} L(\theta, \nu)$

Convexity: $\nabla^2 L(\theta, \nu) = \nabla^2 f_0(\theta) + \nu^T \nabla^2 h(\theta)$

it will depend on the value and sign of ν

(RP) $\begin{cases} \min & f(\theta) + \frac{1}{2}\rho\, h(\theta)^2 \\ \text{s.t.} & h(\theta) = 0 \quad (\nu) \end{cases}$

Lagrangian: $L(\theta, \nu) = f_0(\theta) + \frac{1}{2}\rho\, h(\theta)^2 + \nu^T h(\theta)$

Dual function: $g(\nu) = \min_{\theta} L(\theta, \nu)$

Convexity: $\nabla^2 L(\theta, \nu) = \nabla^2 f_0(\theta) + \underline{\rho \nabla h(\theta)^T \nabla h(\theta)} + \nu^T \nabla^2 h(\theta)$

taking large ρ will be always d.p

* Gradient of the dual function

$g(\nu) = \min_{\theta} f_0(\theta) + \frac{1}{2}\rho\, h(\theta)^2 + \nu^T h(\theta)$

$\theta^*(\nu) = \arg\min_{\theta} f_0(\theta) + \frac{1}{2}\rho\, h(\theta)^2 + \nu^T h(\theta) \Rightarrow \nabla f_0(\theta^*) + [\nu + \rho h(\theta^*)] \nabla h(\theta^*) = 0$

$g(\nu) = f_0(\theta^*(\nu)) + \frac{1}{2}\rho\, h(\theta^*(\nu))^2 + \nu^T h(\theta^*(\nu))$

$\nabla g(\nu) = [\underbrace{\nabla f_0(\theta^*) + [\nu_k + \rho h(\theta^*)] \nabla h(\theta^*)}_{=0}] \nabla \theta^*(\nu) + h(\theta^*(\nu))$

$$\boxed{\nabla g(\nu) = h(\theta^*(\nu))}$$

* <u>Remarks</u>:

a) Since we want to $\{\max_{\nu} g(\nu)\}$, then $\nu_{k+1} = \nu_k + \rho h(\theta_{k+1})$ can be seen as an steepest ascend (gradient) direction where ρ the line-search parameter

b) If we choose ρ s.t we maximize the dual step will make increase the rate convergence of the dual problem

c) The hessian of the dual

$$\nabla^2 g(\nu) = -\nabla h(\theta)\left[\nabla^2 \mathcal{L}(\theta,\nu) + \rho \nabla h(\theta)^T \nabla h(\theta)\right]^{-1} \nabla h(\theta)$$

If $\rho \to \infty \Rightarrow \nabla^2 g(\nu) = \frac{1}{\rho} I$

Thus we approximate the "expensive" $\nabla^2 g(\nu) \simeq \frac{1}{\rho} I$

A preconditioned / Newton iteration is

$$\nu_{k+1} = \nu_k + \left(\frac{1}{\rho} I\right)^{-1} h(\theta^*) = \nu_k + \rho h(\theta^*)$$

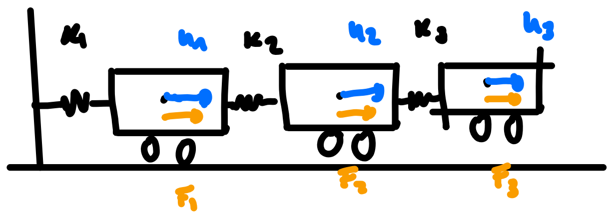
d) if $\rho \to \infty$ Primal: $\operatorname{cond}(\nabla \mathcal{L} + \rho \nabla h^T \nabla h) \to \infty$

if $\rho \to \infty$ Dual: $\operatorname{cond}(\nabla^2 g) \longrightarrow 1$

$\rho \begin{cases} \text{large more importance to the dual} \\ \text{small " " " " primal} \end{cases}$

e) • Interpretation of ρ in mechanics

$$(B)\begin{cases} \min \frac{1}{2} u^T K u - f^T u \\ \text{s.t. } u_1 = 0 \end{cases}$$



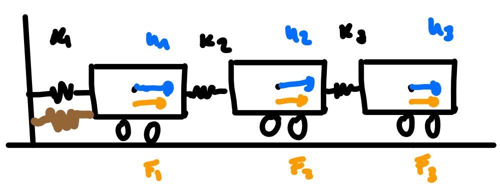
$$K = \begin{bmatrix} k_1+k_2 & -k_2 & 0 \\ -k_2 & k_2+k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}$$

$$F = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$\frac{1}{2} u^T \begin{bmatrix} \rho & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} u$$

$$(RP)\begin{cases} \min \frac{1}{2} u^T K u - f^T u + \frac{1}{2}\rho u_1^2 \\ \text{s.t. } u_1 = 0 \end{cases} \Rightarrow \begin{cases} \min \frac{1}{2} u^T K u - f^T u + \frac{1}{2}\rho u_1^2 \\ \text{s.t. } u_1 = 0 \end{cases}$$

$$\tilde{K} = \begin{bmatrix} k_1+k_2+\rho & -k_2 & 0 \\ -k_2 & k_2+k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}$$



if $\rho \to \infty$

Adjoint problem (finite dimensional)

reduced cost function

* PDE-constrained opt
* Optimal control

$$\begin{cases} \min\limits_{x,y} & J(x,y) \\ \text{s.t.} & A(x)\,y = b(x) \quad (\nu) \end{cases} \xrightarrow[y = A^{-1}(x) b]{\text{with } A \text{ invertible}} \left\{ \min\limits_{x} J(x, A(x)^{-1} b(x)) \right.$$

$$A(x)A^{-1}(x) = I$$
$$\frac{\partial}{\partial x}[A(x)]A^{-1}(x) + A(x)\frac{\partial}{\partial x}[A^{-1}(x)] = 0$$
$$\frac{\partial}{\partial x}[A^{-1}(x)] = -A^{-1}(x)\frac{\partial}{\partial x}[A(x)]A^{-1}(x)$$

a) Gradient computation : $\nabla J = \frac{\partial J}{\partial x} + \frac{\partial J}{\partial y}\frac{\partial y}{\partial x}$; $\frac{\partial y}{\partial x} = \frac{\partial}{\partial x}(A^{-1}(x))b + A^{-1}(x)\frac{\partial b(x)}{\partial x} \overset{\downarrow}{=} -A^{-1}(x)\frac{\partial A(x)}{\partial x}A^{-1}(x)b(x) + A^{-1}(x)\frac{\partial b(x)}{\partial x}$

$$= A^{-1}(x)\left[-\frac{\partial A(x)}{\partial x}A^{-1}(x)b(x) + \frac{\partial b(x)}{\partial x}\right]$$

reduced gradient $\nearrow$ $\nabla_x J = \frac{\partial J}{\partial x} + \underbrace{\frac{\partial J}{\partial y}A^{-1}(x)}_{p}\left[-\frac{\partial A(x)}{\partial x}\underbrace{A^{-1}(x)b}_{y} + \frac{\partial b(x)}{\partial x}\right] = \frac{\partial J}{\partial x} + p\left[-\frac{\partial A(x)}{\partial x}y + \frac{\partial b(x)}{\partial x}\right]$ with $A(x)\,p = \frac{\partial J}{\partial y}$

* Algorithm for reduced cost : Gradient method

$$\begin{cases} (1) \text{ solve : } & A(x)\,y = b \\ (2) \text{ solve : } & A(x)\,p = \frac{\partial J}{\partial y}(x,y) \\ (3) \text{ compute : } & \nabla_x J = \frac{\partial J}{\partial x} + p\left[-\frac{\partial A(x)}{\partial x}y + \frac{\partial b(x)}{\partial x}\right] \\ (4) & x = x - \alpha \nabla_x J \quad \text{with } \alpha \text{ s.t. } J(x - \alpha\nabla_x J) < J(x) \end{cases}$$

b) Interpretation as lagrange multiplier

$$\max_{p}\min_{x,y} \mathcal{L}(x,y,p) = \max_{p}\min_{x,y} J(x,y) + p^T[A(x)\,y - b(x)]$$

KKT
$$\begin{cases} \partial_p \mathcal{L}(x,y,p) = A(x)\,y - b(x) = 0 & \longrightarrow A(x)\,y = b(x) \text{ (fix point)} \\ \partial_y \mathcal{L}(x,y,p) = \frac{\partial J}{\partial y} + p^T A(x) = 0 & \longrightarrow A(x)\,p = \frac{\partial J}{\partial y} \text{ (fix point)} \\ \partial_x \mathcal{L}(x,y,p) = \frac{\partial J}{\partial x} + p^T\left[\frac{\partial A}{\partial x}(x)\,y - \frac{\partial b}{\partial x}(x)\right] = 0 & \longrightarrow \text{(gradient)} \end{cases}$$

* Proposed algorithm :

(same algorithm)

$$\begin{cases} (1) \text{ solve : } & A(x)\,y = b \\ (2) \text{ solve : } & A(x)\,p = \frac{\partial J}{\partial y} \\ (3) \text{ compute : } & \partial_x \mathcal{L} = \frac{\partial J}{\partial x} + p\left[-\frac{\partial A(x)}{\partial x}y + \frac{\partial b}{\partial x}(x)\right] \\ (4) & x = x - \alpha\,\partial_x \mathcal{L} \quad \text{with } \alpha \text{ s.t. } J(x - \alpha\nabla_x J) < J(x) \end{cases}$$

Adjoint problem

$$\left\{ \min_{x} J(x, y(x)) \right\}$$

→ Think constraint as always fulfilled

This is find $y(x)$ solution of the constraint

Example

$a(x,y,v) = \int \nabla^s v \, \phi(x) \nabla^s y$ ← linear with y

$a_x(x,y,v)\tilde{x} = \int \nabla^s v \frac{\partial \phi}{\partial x}(x) \nabla^s y \, \tilde{x}$

$a_y(x,y,v)\tilde{y} = \int \nabla^s v \, \phi(x) \nabla^s \tilde{y} = a(x,\tilde{y},v)$

$\ell(x,v) = \int f(x) v$

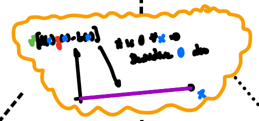
→ Finite dimensional (FE flavour)

$$\begin{cases} \min_{x,y} & J(x,y) \\ \text{s.t.} & v^T A(x) y = v^T b(x) \quad \forall v \end{cases}$$

Gradient computation : $\nabla_x J = \frac{\partial J}{\partial x} + \frac{\partial J}{\partial y}\frac{\partial y}{\partial x}$;

$v^T[A(x)y - b(x)] = 0 \quad \forall v$) Take $y(x)$ st. $\forall x$

$v^T[A(x)y(x) - b(x)] = 0 \quad \forall x \; \forall v$



$v^T\left[\frac{\partial}{\partial x}A(x)y(x) - \frac{\partial b(x)}{\partial x} + A(x)\frac{\partial y(x)}{\partial x}\right] = 0 \quad \forall x \; \forall v$

$v^T A(x) \frac{\partial y(x)}{\partial x} = -v^T\left[\frac{\partial}{\partial x}A(x)y(x) - \frac{\partial b(x)}{\partial x}\right] \quad \forall x \; \forall v$

$-\left(\frac{\partial y}{\partial x}\right)^T A^T(x) v = v^T\left[\frac{\partial}{\partial x}A(x)y - \frac{\partial b(x)}{\partial x}\right] \quad \forall x \; \forall v$

Take $v = p$ st. $A^T(x)p = -\frac{\partial J}{\partial y}$

$\cdot\left(\frac{\partial y}{\partial x}\right)^T A^T(x) p = \left(\frac{\partial y}{\partial x}\right)^T \frac{\partial J}{\partial y} = p^T\left[\frac{\partial}{\partial x}A(x)y - \frac{\partial b(x)}{\partial x}\right] \quad \forall x$

Thus : $\nabla_x J = \frac{\partial J}{\partial x} + \frac{\partial J}{\partial y}\frac{\partial y}{\partial x}$

$= \frac{\partial J}{\partial x} + p^T\left[\frac{\partial b(x)}{\partial x} - \frac{\partial}{\partial x}A(x)y\right] \quad \forall x$

with $A^T(x)p = -\frac{\partial J}{\partial y}$

→ Proposed algorithm : (same algorithm)

(1) solve : $A(x)y = b$

(2) solve : $A(x)p = -\frac{\partial J}{\partial y}$

(3) compute : $\nabla_x J = \frac{\partial J}{\partial x} + p\left[\frac{\partial A(x)}{\partial x}y - \frac{\partial b(x)}{\partial x}\right]$

(4) $x = x - \alpha \nabla_x J$

with α s.t. $J(x - \alpha\nabla_x J) < J(x)$

→ Infinitesimal dimensional case

$$\begin{cases} \min_{x,y} & J(x,y) \\ \text{s.t.} & a(x,y,v) = \ell(x,v) \quad \forall v \end{cases}$$

Differential computation

$DJ(x,y(x))\tilde{x} = D_x J(x,y(x))\tilde{x} + D_y J(x,y(x)) \overbrace{D_x y(x)\tilde{x}}^{\tilde{y}} \quad \forall \tilde{x}$

$a(x,y,v) - \ell(x,v) = 0 \quad \forall v$

$a(x,y(x),v) - \ell(x,v) = 0 \quad \forall v$

← linear

$[a_x(x,y(x),v) - \ell_x(x,v)]\tilde{x} + \overbrace{a_y(x,y(x),v)D_x y(x)\tilde{x}}^{a(x,D_x y(x)\tilde{x},v)} = 0 \quad \forall v \; \forall \tilde{x}$

$-a(x,D_x y(x)\tilde{x},v) = [a_x(x,y(x),v) - \ell_x(x,v)]\tilde{x} \quad \forall v \; \forall \tilde{x}$

Self-adjoint

$-a(x,v,D_x y(x)\tilde{x}) = [a_x(x,y(x),v) - \ell_x(x,v)]\tilde{x} \quad \forall v \; \forall \tilde{x}$

Take $v = p$ st. $a(x,p,D_x y(x)\tilde{x}) = -D_y J(x,y(x)) D_x y(x)\tilde{x} \quad \forall \tilde{x}$

$-a(x,p,D_x y(x)\tilde{x}) = D_y J(x,y(x)) D_x y(x)\tilde{x} = [a_x(x,y(x),p) - \ell_x(x,p)]\tilde{x} \quad \forall \tilde{x}$

Thus :

$DJ(x,y(x))\tilde{x} = D_x J(x,y(x))\tilde{x} + [a_x(x,y(x),p) - \ell_x(x,p)]\tilde{x} \quad \forall \tilde{x}$

with p solution $a(x,p,D_x y(x)\tilde{x}) = -D_y J(x,y(x)) D_x y(x)\tilde{x} \quad \forall \tilde{x}$

Calling $w = D_x y(x)\tilde{x}$ $\quad a(x,p,w) = -D_y J(x,y(x))w \quad \forall w$

→ Proposed algorithm : (same algorithm)

(1) solve : $a(x,y,v) = \ell(x,v) \quad \forall v$

(2) solve : $a(x,p,w) = -D_y J(x,y(x))w \quad \forall w$

(3) compute : $DJ(x,y(x))\tilde{x} = [D_x J(x,y(x)) + a_x(x,y(x),p) - \ell_x(x,p)]\tilde{x}$

$= (g, \tilde{x})_M$ ← metric to be chosen $\begin{cases} H^1, L^2 \\ H^1, M \end{cases}$ many choices

gradient

(4) $x = x - \alpha g$

with α s.t. $J(x - \alpha g) < J(x)$