So, methodology for using Th. (13.2) delivers a set $\tilde{\Lambda}$ such that:

$$\forall \lambda \in \tilde{\Lambda}: \quad L[x(\lambda),\lambda]_x = 0 \;, \quad h(x(\lambda)) = 0$$
$$L[x(\lambda),\lambda]_{xx} = 0$$

$\underset{\boxed{\text{Th }(13.2)}}{\Rightarrow}$ $x(\lambda)$ is a local minimizer for (E).

$\Rightarrow$ $\tilde{X}$ = set of local minimizers

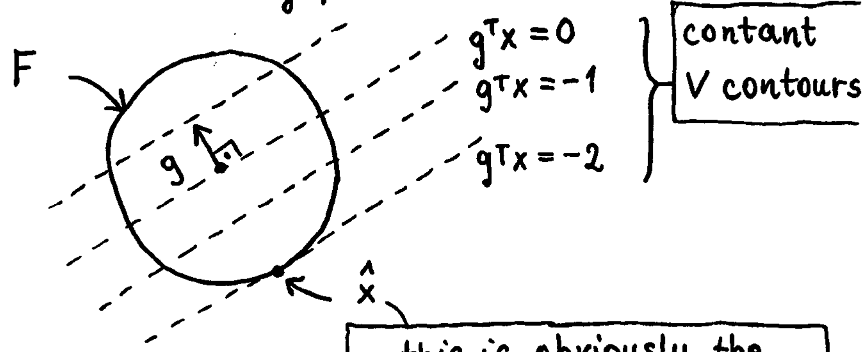
$\Rightarrow$ $x(\lambda) \in \tilde{X}$ yielding least V $\underset{\boxed{\text{hopefully}}}{=}$ global minim.

> in some cases it will not be possible to follow the above methodology exactly
> - but the approach should be applied as far as possible.
>
> Also:
> the methodology is usually practicable, for nonlinear h, only when $m = 1$, i.e. when there is ONE scalar equality constraint.

<u>Example</u>

$$\min_{x \in \mathbb{R}^2} \{ \underbrace{g^T x}_{V(x)} : \underbrace{x_1^2 + x_2^2 - 4}_{h_1(x)} = 0 \}$$

$g \neq 0$



> this is obviously the only local minimizer in this simple example
> - does our methodology find it?

The Lagrangian:

$$L[x,\lambda_1] = V(x) + \lambda_1 h_1(x)$$
$$= g^T x + \lambda_1 (x_1^2 + x_2^2 - 4)$$

(1) choose $x(\lambda_1)$ so $L[x(\lambda_1), \lambda_1]_x = 0$

$$L[x,\lambda]_x = \left[\frac{\partial L}{\partial x_1} \quad \frac{\partial L}{\partial x_2} \right]$$

$$= \left[g_1 + 2\lambda_1 x_1 \quad g_2 + 2\lambda_1 x_2 \right] = 0$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\frac{1}{2\lambda_1} \underbrace{\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}}_{x(\lambda_1)}$$

$$\Rightarrow x(\lambda_1) = -\frac{1}{2\lambda_1} g$$

(2) find $\Lambda = \{ \lambda_1 : h_1[x(\lambda_1)] = 0 \}$

$$h_1[x(\lambda_1)] = x_1(\lambda_1)^2 + x_2(\lambda_1)^2 - 4$$

$$= \| x(\lambda_1) \|^2 - 4 = \frac{1}{4\lambda_1^2} \|g\|^2 - 4$$

$$= 0 \quad \text{iff} \quad \lambda_1 = \pm \frac{\|g\|}{4}$$

so

$$\Lambda = \{ \underbrace{-\tfrac{1}{4}\|g\|}_{\bar{\lambda}_1}, \underbrace{+\tfrac{1}{4}\|g\|}_{\tilde{\lambda}_1} \}$$

(3) find $\tilde{\Lambda} = \{ \lambda_1 \in \Lambda : L[x(\lambda_1), \lambda_1]_{xx} > 0 \}$

$$L[x, \lambda_1] = g^T x + \lambda_1 (x_1^2 + x_2^2 - 4)$$

$$\Rightarrow L[x,\lambda_1]_{xx} = \begin{bmatrix} 2\lambda_1 & 0 \\ 0 & 2\lambda_1 \end{bmatrix}$$

Therefore:

- $L[x(\bar{\lambda}_1), \bar{\lambda}_1]_{xx} = -\frac{1}{2}\|g\| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \not> 0$ (#)

  $(\Rightarrow \bar{\lambda}_1 \notin \tilde{\Lambda})$

- $L[x(\tilde{\lambda}_1), \tilde{\lambda}_1]_{xx} = \frac{1}{2}\|g\| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} > 0$ (#)

  $(\Rightarrow \tilde{\lambda}_1 \in \tilde{\Lambda})$

Hence:

$$\tilde{\Lambda} = \{ \tilde{\lambda}_1 \} = \{ \tfrac{1}{4}\|g\| \}.$$

---

a diagonal matrix is positive definite iff all its diagonal entries are >0.

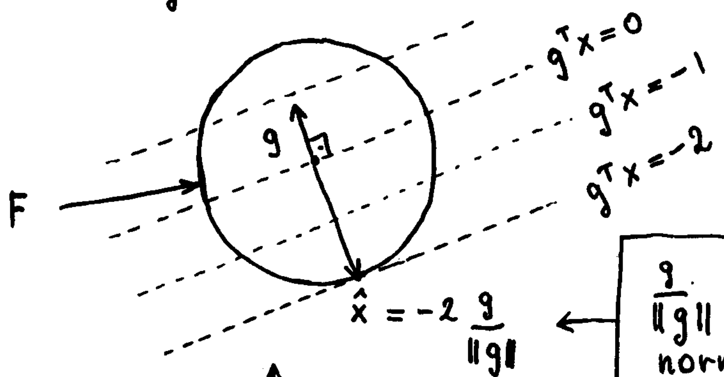
④ find $\tilde{X} = \{x(\lambda_1) : \lambda_1 \in \tilde{\Lambda}\}$

$= \{x(\tilde{\lambda}_1) = \{-2\frac{g}{\|g\|}\}$

$=$ our set of local minimizers

So: $\hat{x} = -\frac{2g}{\|g\|}$

For this very simple case, we can check geometrically:



$\frac{g}{\|g\|}$ is the normalized vector g

we see from this diagram that our local minimizer $\hat{x}$ is actually the global minimizer ← generally it is not always so nice.

Equality constrained optimization via duality

In duality theory, the original opt. problem is called the primal problem (P):

Our primal problem is:

$$\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = 0\} \qquad (P)$$

$\boxed{h : \mathbb{R}^n \to \mathbb{R}^m}$

We shall derive an "equivalent" dual problem (D), potentially easier to solve than (P), from which the solution to (P) can be found easily.

Solution of (P) will be aided by study of

$$\min_{x \in \mathbb{R}^n} \{V(x) : h(x) = y\}$$

$\in \mathbb{R}^n$

a generalized problem owing to y

Define :

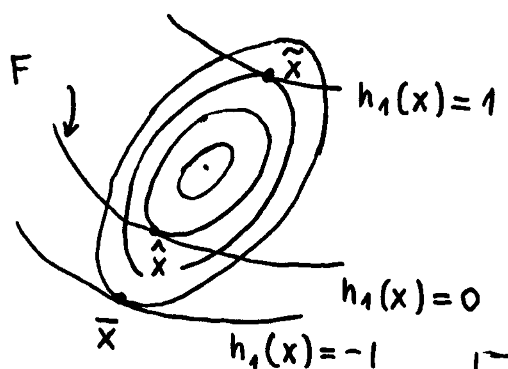
- $\pi(y) \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : \underbrace{h(x)=y}_{\substack{\text{perturbation of}\\ \text{primal constraint}\\ h(x)=0}} \}$

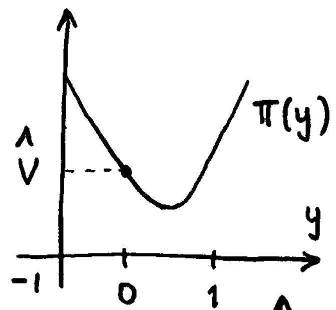
  the perturbation function

- $\hat{V} \triangleq \min_{x \in \mathbb{R}^n} \{ V(x) : h(x) = 0 \}$

  = the minimal cost for (P).

(14.1) Th $\quad \pi(0) = \hat{V} \leftarrow \boxed{\text{obvious}}$





this kind of plot is the key to duality theory

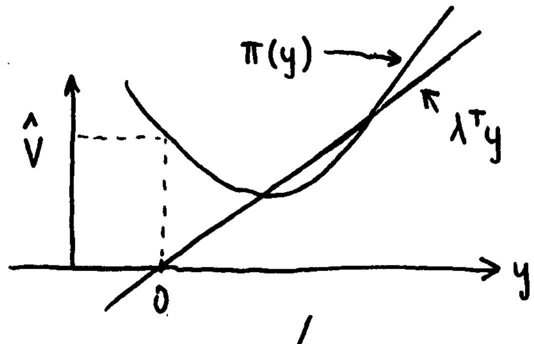
$\pi(-1) = V(\bar{x})$
$\pi(0) = V(\hat{x}) = \hat{V}$
$\pi(1) = V(\tilde{x})$

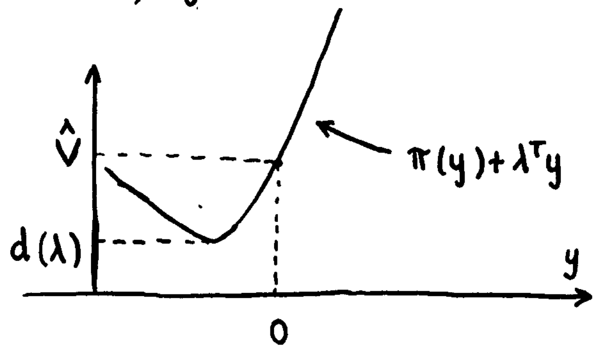
Duality theory relies heavily on the dual function:

$$\underbrace{d(\lambda)}_{\in \mathbb{R}^m} \triangleq \min_{y \in \mathbb{R}^m} [\pi(y) + \lambda^T y] \qquad (DF)$$

graphically :

adding $\lambda^T y$ "rotates" π about $y=0$





(14.2) Th

(i) $\pi(y) + \lambda^T y \,|_{y=0} = \hat{V}, \quad \forall \lambda \in \mathbb{R}^m$

(ii) $d(\lambda) \leq \hat{V}, \quad \forall \lambda \in \mathbb{R}^m$

obvious

Proof of Th (14.2) (ii):

$$d(\lambda) = \min_{y \in \mathbb{R}^m} \left[\pi(y) + \lambda^T y\right]$$

$$\leq \left[\pi(y) + \lambda^T y\right]_{|y=0}$$

$$= \pi(0) = \hat{V} \quad , \quad \forall \lambda \in \mathbb{R}^m$$

QED.

Evaluation of $d(\lambda)$:

$$d(\lambda) = \min_{y \in \mathbb{R}^m} \left[\underbrace{\pi(y)}_{= \min\limits_{x \in \mathbb{R}^n} \{V(x) : h(x) = y\}} + \lambda^T y\right]$$

$\Downarrow$

$\boxed{\text{finding } d(\lambda) \text{ this way is at least as hard as solving (P) so, this way seems useless as an aid to solving (P)}}$

BUT ... saved by $\downarrow$

(14.3) Th $d(\lambda)$ is also given by

$$d(\lambda) = \min_{x \in \mathbb{R}^n} \left[V(x) + \lambda^T h(x)\right]$$

$\boxed{\text{unconstrained minimization relatively easy to do using secant alg. etc.}}$

Proof:

(14.3) Let $\mathcal{J}(x,y) \triangleq \begin{cases} V(x) & \text{if } h(x) = y \\ +\infty & \text{if } h(x) \neq y \end{cases}$

$$\theta \triangleq \min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \left\{\mathcal{J}(x,y) + \lambda^T y\right\}$$

Then:

$$\theta = \min_{x \in \mathbb{R}^n} \left[\min_{y \in \mathbb{R}^m} \left\{\mathcal{J}(x,y) + \lambda^T y\right\}\right] \quad (\#)$$

$$= \min_{y \in \mathbb{R}^m} \left[\min_{x \in \mathbb{R}^n} \left\{\mathcal{J}(x,y) + \lambda^T y\right\}\right] \quad (£)$$

From (#):

$$\theta = \min_{x \in \mathbb{R}^n} \left[\min_{y \in \mathbb{R}^m} \{ J(x,y) + \lambda^T y \} \right]$$

$$= \min_{x \in \mathbb{R}^n} \left[J(x,\hat{y}) + \lambda^T \hat{y} \,\middle|\, \hat{y} = h(x) \right]$$

$$\boxed{\begin{array}{l} \text{as } J(x,y) + \lambda^T y = \infty \\ \text{if } h(x) \neq y \end{array}}$$

$$= \min_{x \in \mathbb{R}^n} \{ \underbrace{J(x,h(x))}_{\boxed{(14.3)}} + \lambda^T h(x) \}$$

$$= \min_{x \in \mathbb{R}^n} \{ V(x) + \lambda^T h(x) \}$$

$$= d(\lambda) \quad \text{of Th (14.3)}$$

From (£);

$$\theta = \min_{y \in \mathbb{R}^m} \left[\min_{x \in \mathbb{R}^n} \{ J(x,y) + \lambda^T y \} \right]$$

$$\Rightarrow \theta = \min_{y \in \mathbb{R}^m} \left[\min_{x \in \{x \in \mathbb{R}^n : h(x) = y\}} \{ J(x,y) + \lambda^T y \} \right]$$

$$\boxed{\begin{array}{l} \text{since } J(x,y) + \lambda^T y = +\infty \\ \text{if } h(x) \neq y \end{array}}$$

$$= \min_{y \in \mathbb{R}^m} \left[\min_{x \in \{x \in \mathbb{R}^n : h(x) = y\}} \{ J(x,h(x)) + \lambda^T y \} \right]$$

$$= \min_{y \in \mathbb{R}^m} \left[\min_{x \in \{x \in \mathbb{R}^n : h(x) = y\}} \{ V(x) + \lambda^T y \} \right]$$

$$= \min_{y \in \mathbb{R}^m} \left[\underbrace{\min_{x \in \{x \in \mathbb{R}^n : h(x) = y\}} \{ V(x) \}}_{=\pi(y)} + \lambda^T y \right]$$

$$= \min_{y \in \mathbb{R}^m} \{ \pi(y) + \lambda^T y \}$$

$$= d(\lambda) \quad \text{of Definition (DF)}$$

Hence:

$$\min_{x\in\mathbb{R}^n}\left\{V(x)+\lambda^T h(x)\right\} \overset{(\#)}{=} \Theta \overset{(£)}{=} \min_{y\in\mathbb{R}^m}\left\{\pi(y)+\lambda^T y\right\}$$

$$\min_{x\in\mathbb{R}^n}\left\{V(x)+\lambda^T h(x)\right\} = d(\lambda) \text{ of Th 14.3}$$

$$\min_{y\in\mathbb{R}^m}\left\{\pi(y)+\lambda^T y\right\} = d(\lambda) \text{ of Def}^n\ (DF)$$

$d(\lambda)$ of Th 14.3 = $d(\lambda)$ of Def^n (DF): same thing.

QED

$d(\lambda)$ of Th 14.3: expression for $d(\lambda)$ which is easy to evaluate

$d(\lambda)$ of Def^n (DF): expression for $d(\lambda)$ which has conceptual importance in constrained optimization