<u>In practice</u>:

except for V = quadratic & h = affine using procedure (15.6) is harder than it seems since we cannot get a nice (*) formula for $d(\lambda)$

$\Rightarrow$ cannot easily find $\hat{\lambda}$ in this way

Use of a "trick" largely overcomes this problem, and also often makes things work for non-convex (P) as well.

<u>The trick</u> is to modify π to π_r $(r > 0)$:

$$\pi_r(y) \triangleq \pi(y) + \frac{r}{2}\|y\|^2 \quad \longleftarrow \boxed{\text{augmented perturbation function}}$$

---

* By a nice formula for $d(\lambda)$ we mean one which can be maximized analytically.
Our formula for $d(\lambda)$ yields a value for $d(\lambda)$ relatively easily but is not easy to maximize analytically.

Then, for big enough r

$$\pi_r \underset{\uparrow}{\cong} \text{approximately quadratic in } y$$

$\boxed{\text{with nice computational consequences}}$

Substituting π by π_r, we can re-do the "duality theory" using:

$$d_r(\lambda) \triangleq \min_{\lambda \in \mathbb{R}^m} \{ \pi_r(y) + \lambda^T y \} \quad \longleftarrow \boxed{\text{augmented dual function}}$$

for which the easily-evaluatable formula is:

$$d_r(\lambda) = \min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \lambda^T h(x) \right\}$$

Then, we get:

$$\hat{V} = \max_{\lambda \in \mathbb{R}^m} d_r(\lambda) \quad \text{with} \quad \hat{\lambda} \underset{\boxed{\text{independent of } r}}{\overset{\boxed{\text{still}}}{=}} -\nabla\pi(0)$$

and:

$$\hat{x} = \tilde{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \hat{\lambda}^T h(x) \right\}$$

for primal problem (P)

Nothing much has changed so far ...
so what is the point of making the augmentations?

The important thing is that:

which justifies using Π_r

$$\hat{\lambda} \cong \lambda + r\,h(\bar{x}), \quad \forall \lambda \qquad (\$)$$

depends on λ

with approximation error decreasing as r increases

where:

$$\bar{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2}\|h(x)\|^2 + \lambda^T h(x) \right\}$$

<u>Amazing consequence:</u>

($) provides a way to find $\hat{\lambda}$ accurately (if r is big enough) <u>without</u> us needing a nice formula for $d(\lambda)$ (or for $d_r(\lambda)$).

<u>Motivation for ($):</u>

(making the "singleton assumption" and using the approach employed in the proof of Th (14.3))

Consider:

$$\min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \left\{ \mathcal{J}(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (\#)$$

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

Call the minimizers $\bar{x}$ & $\bar{y}$!

$$= \min_{x \in \mathbb{R}^n} \min_{y \in \mathbb{R}^m} \left\{ \mathcal{J}(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

for each x, the minimizing y is $\hat{y}(x) = h(x)$ (£) as in the proof of Th. (14.3)

$= \dots$

$$\dots = \min_{x \in \mathbb{R}^n} \left\{ V(x) + \frac{r}{2} \|h(x)\|^2 + \lambda^T h(x) \right\} \qquad (¥)$$

the minimizer here must be the minimizer $\bar{x}$ for (#) (and vice versa). So, from (£), the minimizer $\bar{y}$ for (#) must be $\bar{y} = \hat{y}(\bar{x}) = h(\bar{x})$.

Summarizing:

- the minimizer $\bar{x}$ for (#) $\equiv$ $\equiv$ the minimizer x for (¥)
- the minimizer $\bar{y}$ for (#) $\equiv$ $\equiv h(\bar{x})$ where $\bar{x}$ can be found by minimizing (¥).

Also, going back to (#) again:

$$\min_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} \left\{ J(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

$$= \min_{y \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \left\{ J(x,y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\}$$

$$\underset{\text{see Th (14.3}}{=} \min_{y \in \mathbb{R}^m} \left\{ \pi(y) + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (£)$$

the minimizer here must be the minimizer $\bar{y} = h(\bar{x})$ for (#)

$$\cong \min_{y \in \mathbb{R}^m} \left\{ \underbrace{\pi(0) + \nabla\pi(0)^T (y-0)}_{\text{1-st-order expansion for } \pi(y)} + \frac{r}{2}\|y\|^2 + \lambda^T y \right\} \qquad (\beta)$$

1-st-order expansion for $\pi(y)$ with approximation error swamped by $\frac{r}{2}\|y\|^2$ if r is big enough.

a quadratic in y, for which the minimizing y is $-\frac{1}{r}[\nabla\pi(0) + \lambda]$

Hence:

$$-\frac{1}{r}\left[\nabla\pi(0)+\lambda\right] \cong \bar{y} = h(\bar{x})$$

> because we argue that, since opt. problems (α) and (β) are approximately the same, their minimizers $\bar{y} = h(\bar{x})$ and $-\frac{1}{r}[\nabla\pi(0)+\lambda]$ are approximately equal.

That is:

$$-\nabla\pi(0) \cong \lambda + rh(\bar{x})$$

> with approximation error decreasing as r increases

i.e. $\hat{\lambda} \cong \lambda + rh(\bar{x})$

> depends on λ

> with approximation error decreasing as r increases

which is (\$), as required.

Now, we can state "the multiplier algorithm" for solving: $\min_{x\in\mathbb{R}^n} \{V(x) : h(x) = 0\}$

and based on duality but does not need a nice formula for $d(\lambda)$.

The Multiplier Algorithm

Choose scalars $r_0 > 0$, $\Psi > 1$, $\lambda_0 \in \mathbb{R}^m$

Set $j=0$

1) [At iteration j] Find

$$\bar{x}_j = \arg\min_{x\in\mathbb{R}^n} \left\{ V(x) + \frac{r_j}{2}\|h(x)\|^2 + \lambda_j^T h(x) \right\}$$

> using a standard alg. for unconstrained optimization

2) Set $\lambda_{j+1} := \lambda_j + r_j\, h(\bar{x}_j)$

> from (\$): this makes λ_{j+1} an approx. to $\hat{\lambda}$ which improves as j increases since r_j increases with j

3) Set $r_{j+1} = \psi r_j$

to increase r_j

4) Set $j := j+1$ and go to 1)

<u>Remark:</u>

If (P) $\neq$ too nasty

- $\lambda_j \longrightarrow \lambda$ $\leftarrow$ owing to ($) and r_j increasing with j
- $\bar{x}_j \longrightarrow \hat{x}$ for (P) $\leftarrow$ because $\lambda_j \to \lambda$

Stop iterating when:

(A) $\| h(\bar{x}_j) \| < \varepsilon_1$ $\Rightarrow$ an adequate approxim. to satisfaction of $h(x)=0$ has been achieved

(B) $\| \bar{x}_j - \bar{x}_{j-1} \| < \varepsilon_2$ $\Rightarrow$ no significant change during last iteration