Majorization. Minimization.

(what its relation to ADMM ?)

① surrogate function: $\bar{F}(x|x^{(k)})$ ($\bar{F}$ majorize F)

s.t. $F(x) \le \bar{F}(x|x^{(k)})$ and. $\bar{F}(x^{(k)}|x^{(k)}) = F(x^{(k)})$

Globally. tangent at $x^{(k)}$

- then we optimize $\bar{F}(x|x^{(k)})$ gets $x^{(k+1)}$

$$\bar{F}(x^{(k+1)}|x^{(k)}) \le \bar{F}(x^{(k)}|x^{(k)})$$

so we have $F(x^{(k+1)}) \le \bar{F}(x^{(k+1)}|x^{(k)}) \le \bar{F}(x^{(k)}|x^{(k)}) = F(x^{(k)})$

$x^{(k+1)}$ also makes $F(x)$ smaller.

so we could optimize the surrogate fcn instead of the original one.

②. divide problem using surrogate fcn.

→ to achieve parallel optimization without inter-device communication.

(if) $\bar{F}(x|x^{(k)}) \approx \sum_{\alpha \in S} \bar{F}^{\alpha}(x^{\alpha}|x^{(k)})$. $\alpha \in S$.

③. ease of optimization of the surrogate fcn.

~~separation~~ of parameters

separation

④ speed of convergence

- general Newton-Raphson update: (second-order)

$$\theta^{(k+1)} = \theta^{(k)} - \left[\nabla^2 \bar{F}(\theta^{(k)}|\theta^{(k)})\right]^{-1} \nabla \bar{F}(\theta^{(k)}|\theta^{(k)})$$

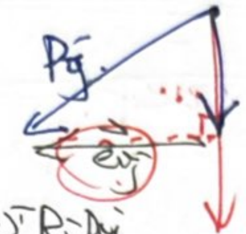
- well-designed $\bar{F}$ (first-order): tends to require more iterations. but simpler iterations. ⇒ general faster convergence.

⑤ Handling Constraints.

- given the initial point. satisfying the constraints
- design surrogate function. confirm the satisfaction of constraints.

Design surrogate function for BA problem

① $d_{ij} p_{ij} = R_i^T (l_j - t_i)$ …… • pose (camera to world) is variable
• pixel seen in image.



↓ $\lambda_{ij} = 1/d_{ij}$

• $p_{ij} - \lambda_{ij} R_i^T (l_j - t_i) = 0$.

• $\lambda_{ij}^* = \arg\min_{\lambda \in R} \| p_{ij} - \lambda_{ij} R_i^T (l_j - t_i) \|^2 = \frac{(l_j - t_i)^T R_i p_{ij}}{\|l_j - t_i\|^2} \sim \frac{[R_i^T (l_j - t_i)]^T p_{ij}}{\|n_i\|^2}$

(dot product of 2 vectors)

$\therefore e_{ij} = \left(I - \frac{R_i^T (l_j - t_i)(l_j - t_i)^T R_i}{\|l_j - t_i\|^2}\right) p_{ij} \in R^3$ ⇒ point to line distance.
(on the normalised sphere)

② make. $P_{ij}(c_i | x^{(k)})$ and $Q_{ij}(l_j | x^{(k)})$ (see paper for details)
(separates the camera and landmark parameters)
s.t. $F_{ij}(c_i, l_j) \le P_{ij}(c_i | x^{(k)}) + Q_{ij}(l_j | x^{(k)})$
satisfies the surrogate condition. (skip details)

③ $F(x) = \sum_{(i,j) \in \varepsilon'} F_{ij}(c_i, l_j) + \sum_{(i,j) \in \varepsilon''} F_{ij}(c_i, l_j)$

((i,j) in same device.) ((i,j) in different devices)
need to handle distribution problem.

$\le \sum_{(i,j) \in \varepsilon'} F_{ij}(c_i, l_j) + \sum_{(i,j) \in \varepsilon''} \left[P_{ij}(c_i | x^{(k)}) + Q_{ij}(l_j | x^{(k)}) \right]$

• camera/point decoupled.
• then could be separated into different devices.

• $G(x | x^{(k)}) \triangleq \sum_{\alpha \in S} G^\alpha (x^\alpha | x^{(k)})$

$= \underbrace{\frac{1}{2} \xi \sum_{\alpha \in S} \| x^\alpha - x^{\alpha(k)} \|^2}_{(\xi > 0,\ \xi > 0,\ \text{for convergence analysis})} + \underbrace{\sum_{(i,j) \in \varepsilon'} F_{ij}(c_i, l_j) + \sum_{(i,j) \in \varepsilon''} \left[P_{ij}(c_i | x^{(k)}) + Q_{ij}(l_j | x^{(k)}) \right]}_{\ge F(x)}$

(satisfies surrogate conditions.)
⇒ majorization minimization.

Nesterov's accelerated methods

- extrapolate the iterate $x^{\alpha(k)}$ with momentum $x^{\alpha(k)} - x^{\alpha(k-1)}$

→ $\bar{x}^{\alpha(k)} = \mathrm{Proj}(x^{\alpha(k)} + \gamma^{\alpha(k)}(x^{\alpha(k)} - x^{\alpha(k-1)}))$

(project to manifolds) (GR ratio of momentum).

(update with fn (skip) see the paper)

- for BA case:

$$\begin{cases} \bar{R}_v^{(k)} = \mathrm{Proj}_{Rot3D}\left(R_v^{(k)} + \gamma^{\alpha(k)}(R_v^{(k)} - R_v^{(k-1)})\right) \\ \bar{t}_v^{(k)} = t_v^{(k)} + \gamma^{\alpha(k)}(t_v^{(k)} - t_v^{(k-1)}) \\ \dots \text{(skip)} \dots \end{cases}$$

- $x^{\alpha(k+1)} \leftarrow \arg\min_{x^\alpha} E^\alpha(x^\alpha | \bar{x}^{(k)})$ for $\alpha \in S$. each device

(difference from the fn before)

(remains decentralized)

- N-a. has no provable convergence.

Adaptive Restart ← solved by

- If $F(x^{(k)})$ not improved by $x^{\alpha(k)}$
- continuous with $x^{(k+1)}$ (without NA)

problem : evaluate the final loss $F(x^{(k)})$ without central device.
recursively update on each device.

→ adaptive restart metrics (skip).